

## “Super-deblended” dust emission in galaxies. I. The GOODS-North catalog and the cosmic star formation rate density out to redshift 6

Article (Published Version)

Sargent, Mark (2018) “Super-deblended” dust emission in galaxies. I. The GOODS-North catalog and the cosmic star formation rate density out to redshift 6. *Astrophysical Journal*, 853 (2). ISSN 0004-637X

This version is available from Sussex Research Online: <http://sro.sussex.ac.uk/id/eprint/73714/>

This document is made available in accordance with publisher policies and may differ from the published version or from the version of record. If you wish to cite this item you are advised to consult the publisher’s version. Please see the URL above for details on accessing the published version.

### **Copyright and reuse:**

Sussex Research Online is a digital repository of the research output of the University.

Copyright and all moral rights to the version of the paper presented here belong to the individual author(s) and/or other copyright owners. To the extent reasonable and practicable, the material made available in SRO has been checked for eligibility before being made available.

Copies of full text items generally can be reproduced, displayed or performed and given to third parties in any format or medium for personal research or study, educational, or not-for-profit purposes without prior permission or charge, provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.



# “Super-deblended” Dust Emission in Galaxies. I. The GOODS-North Catalog and the Cosmic Star Formation Rate Density out to Redshift 6

Daizhong Liu<sup>1,2,3,4</sup> , Emanuele Daddi<sup>2</sup> , Mark Dickinson<sup>5</sup>, Frazer Owen<sup>6</sup> , Maurilio Pannella<sup>7</sup> , Mark Sargent<sup>8</sup> ,  
Matthieu Béthermin<sup>9,10</sup> , Georgios Magdis<sup>11</sup> , Yu Gao<sup>1</sup> , Xinwen Shu<sup>2,12</sup> , Tao Wang<sup>2</sup> , Shuowen Jin<sup>2</sup> , and Hanae Inami<sup>13</sup>

<sup>1</sup> Purple Mountain Observatory/Key Laboratory for Radio Astronomy, Chinese Academy of Sciences,

No. 2 West Beijing Road, Nanjing, People’s Republic of China; [dzliu@mpia.de](mailto:dzliu@mpia.de)

<sup>2</sup> CEA, IRFU, DAP, AIM, Université Paris-Saclay, Université Paris Diderot, Sorbonne Paris Cité, CNRS, F-91191 Gif-sur-Yvette, France

<sup>3</sup> University of Chinese Academy of Sciences, 19A Yuquan Road, Shijingshan District, 10049, Beijing, People’s Republic of China

<sup>4</sup> Max Planck Institute for Astronomy, Königstuhl 17, D-69117 Heidelberg, Germany

<sup>5</sup> National Optical Astronomy Observatory, Tucson, AZ 85719, USA

<sup>6</sup> National Radio Astronomy Observatory, Socorro, NM 87801, USA

<sup>7</sup> Faculty of Physics, Ludwig-Maximilians Universität, Scheinerstr. 1, D-81679 Munich, Germany

<sup>8</sup> Astronomy Centre, Department of Physics and Astronomy, University of Sussex, Brighton BN1 9QH, UK

<sup>9</sup> European Southern Observatory, Karl Schwarzschild Strasse 2, D-85748, Garching, Germany

<sup>10</sup> Aix Marseille Univ, CNRS, LAM, Laboratoire d’Astrophysique de Marseille, Marseille, France

<sup>11</sup> Dark Cosmology Centre, Niels Bohr Institute, University of Copenhagen, Juliane Mariesvej 30, DK-2100 Copenhagen, Denmark

<sup>12</sup> Anhui Normal University, Wuhu, Anhui, People’s Republic of China

<sup>13</sup> Observatoire de Lyon, 9, Avenue Charles André, F-69561 Saint Genis Laval, France

Received 2017 March 13; revised 2017 December 28; accepted 2018 January 2; published 2018 February 5

## Abstract

We present a new technique to measure multi-wavelength “super-deblended” photometry from highly confused images, which we apply to *Herschel* and ground-based far-infrared (FIR) and (sub-)millimeter (mm) data in the northern field of the Great Observatories Origins Deep Survey. There are two key novelties. First, starting with a large database of deep *Spitzer* 24  $\mu\text{m}$  and VLA 20 cm detections that are used to define prior positions for fitting the FIR/submm data, we perform an *active* selection of *useful* priors independently at each frequency band, moving from less to more confused bands. Exploiting knowledge of redshift and all available photometry, we identify *hopelessly faint* priors that we remove from the fitting pool. This approach significantly reduces blending degeneracies and allows reliable photometry to be obtained for galaxies in FIR+mm bands. Second, we obtain well-behaved, nearly Gaussian flux density uncertainties, individually tailored to all fitted priors for each band. This is done by exploiting extensive simulations that allow us to calibrate the conversion of formal fitting uncertainties to realistic uncertainties, depending on directly measurable quantities. We achieve deeper detection limits with high fidelity measurements and uncertainties at FIR+mm bands. As an illustration of the utility of these measurements, we identify 70 galaxies with  $z \geq 3$  and reliable FIR+mm detections. We present new constraints on the cosmic star formation rate density at  $3 < z < 6$ , finding a significant contribution from  $z \geq 3$  dusty galaxies that are missed by optical-to-near-infrared color selection. Photometric measurements for 3306 priors, including more than 1000 FIR+mm detections, are released publicly with our catalog.

**Key words:** galaxies: ISM – galaxies: photometry – galaxies: star formation – infrared: galaxies – techniques: photometric

**Supporting material:** machine-readable table

## 1. Introduction

A wealth of infrared (IR) to millimeter (mm) deep survey observations has accumulated in the last decade, with data obtained by the *Spitzer Space Telescope* (hereafter *Spitzer*; Werner et al. 2004), *Herschel Space Observatory* (hereafter *Herschel*; Pilbratt et al. 2010), and many ground-based single-dish telescopes (e.g., the IRAM 30 m and JCMT 15 m telescopes). These observations are indispensable for understanding the evolution of star formation and of the interstellar medium of galaxies from early cosmic epochs to the present. Dust grains are produced on short timescales during the evolution of massive stars, and hence are a direct product of star formation activity in young star-forming galaxies.<sup>14</sup> Dust absorbs starlight at ultra-violet (UV) to optical wavelengths and

re-emits the energy as longer-wavelength IR photons. Large amounts of dust can exist in galaxies with intense star formation activity, strongly attenuating their rest-frame UV-to-optical light. Hence, with UV to optical observations alone, one can directly measure only relatively unobscured star formation in galaxies. The dust attenuation can be estimated from the UV to optical photometry, for example from the UV continuum slope or the UV-to-optical spectral energy distribution (SED), but these estimates are indirect, they can have large uncertainties, and they can entirely fail to detect large amounts of star formation hiding behind optically thick dust. Far-IR (FIR) observations have thus become an essential tool for directly measuring obscured star formation rates (SFRs) in galaxies across cosmic time (e.g., Draine et al. 2007; Magdis et al. 2012; Ciesla et al. 2014; Tan et al. 2014).

*Spitzer* was the first IR space telescope efficient and sensitive enough to survey large areas with adequate sensitivity to detect galaxies at cosmological redshifts (e.g., Le Floch et al. 2005;

<sup>14</sup> Dust grains can also be produced by low- and intermediate-mass stars during their Asymptotic Giant Branch (AGB) stage, on longer timescales—for example, Draine (2003), Gail et al. (2009), and Zhukovska & Henning (2013).

Frazer et al. 2006, 2009; Le Floch et al. 2009; Magnelli et al. 2009, 2011; Ashby et al. 2013, 2015). *Spitzer* observations are very sensitive at 3.6–8.0  $\mu\text{m}$  (the Infrared Array Camera, IRAC; Fazio et al. 2004) and at 24  $\mu\text{m}$  (the Multiband Imaging Photometer, MIPS; Rieke et al. 2004), but is much less sensitive at FIR wavelengths (70 and 160  $\mu\text{m}$ ) than the later IR space telescope *Herschel*.

*Herschel* observed at FIR wavelengths from 70 to 500  $\mu\text{m}$ . Its Photoconductor Array Camera and Spectrometer (PACS; Poglitsch et al. 2010) provided more than four times better angular resolution than *Spitzer* at 70–160  $\mu\text{m}$ , while the Spectral and Photometric Imaging Receiver (SPIRE; Griffin et al. 2010) covered longer wavelengths (250, 350 and 500  $\mu\text{m}$ ) at very efficient mapping speeds. *Herschel* enabled direct measurements of FIR emission for a large number of galaxies at cosmological redshifts (e.g., Gruppioni et al. 2013; Lee et al. 2013; Magnelli et al. 2013; Béthermin et al. 2015; Schreiber et al. 2015; Shu et al. 2016; Wang et al. 2016).

However, source confusion problems can introduce substantial biases in photometric works. *Herschel* SPIRE images have point-spread functions (PSFs) that are several times larger (17''6–35''2) than those of PACS images (7''–12''). The fluxes of individual galaxies are often difficult to measure with SPIRE because of blending from close neighbors. If one considers all possible star-forming galaxies that might contribute to the SPIRE signal, and try to simultaneously fit for all of them, no individual measurement can be obtained because of degeneracies. In sufficiently deep data, the average number of sources that *potentially* contribute within each SPIRE beam is always  $\gg 1$ . On the other hand, measured fluxes will be biased if we simply ignore a fraction of the potential contributing sources in the *prior-extraction* method,<sup>15</sup> or if we treat several blended sources as one source (e.g., the *blind-extraction* method<sup>16</sup>). Therefore, a careful approach is needed to preselect the sources used as fitting priors so that the actual number of fitted sources is  $\lesssim 1$  per beam for all bands. In addition, the flux contribution of faint sources that are *not* included in the fitting should be taken into account.

Several solutions have been explored for prior-extraction of galaxy fluxes in FIR/mm images with potentially significant source confusion. Béthermin et al. (2010b) started from an input catalog of MIPS 24  $\mu\text{m}$  source positions (Béthermin et al. 2010a) and applied an IDL-based routine with a linear inversion algorithm (FASTPHOT) to perform prior-extraction photometry for BLAST data in the Extended *Chandra* Deep Field South. Roseboom et al. (2010) used MIPS 24  $\mu\text{m}$  detected sources ( $S/N \geq 5$ ) as priors for SPIRE photometry with their cross-identification (XID) tool.<sup>17</sup> Elbaz et al. (2011)<sup>18</sup> also use MIPS 24  $\mu\text{m}$  detected ( $S/N \geq 5$ ) sources for PACS 160  $\mu\text{m}$  and SPIRE 250  $\mu\text{m}$  photometry in GOODS-North and GOODS-South fields, adopting SPIRE 250  $\mu\text{m}$   $S/N > 2$  sources as the prior list for SPIRE 350 and 500  $\mu\text{m}$  photometry. Béthermin et al. (2012, 2015) used MIPS 24  $\mu\text{m}$  sources as priors as well, but have done more detailed selections. They first perform

stacking to derive an average relationship between redshift, stacked flux and SPIRE-to-24  $\mu\text{m}$  color (flux ratio), based on which they predict the SPIRE flux for each prior source. A 24  $\mu\text{m}$  source is selected as a prior only if it has the highest predicted SPIRE flux within  $0.5 \times \text{PSF FWHM}$ , while the remaining fainter sources are ignored. Lee et al. (2013) make use not only of MIPS 24  $\mu\text{m}$  but also VLA 1.4 GHz sources as priors for the SPIRE photometry in the COSMOS field. They use the XID algorithm and the iterative jackknife approach for the photometry as also done by Roseboom et al. (2010, 2012). Yan et al. (2014) presented a careful deblending work toward a few specific sources. Their approach uses *H*-band prior sources to form different decomposition schemes, and runs iteratively to determine a best fitting scheme.

Safarzadeh et al. (2015) developed a novel technique based on Markov chain Monte Carlo (MCMC) sampling for prior-extraction photometry. Their approach uses *H*-band sources as position priors, and fits UV-to-optical SEDs to predict FIR flux densities (accurate within  $\pm 1$  dex; these serve as the initial guesses for MCMC fitting). Then, they separate all priors into blended groups (so as to save computational resources) and fit each blended group in a MCMC way to obtain the posterior probability distribution function (PDF). They tested their technique with a simulated PACS 160  $\mu\text{m}$  image and demonstrated such an approach can obtain reliable fluxes below the nominal confusion limit. Hurley et al. (2017) developed a similar MCMC-based prior-extraction tool XID<sup>+</sup>, which can fit all prior sources to obtain the posterior PDF of the flux of each prior source. Then they measured SPIRE fluxes for COSMOS 24  $\mu\text{m}$  prior sources, tiling the full image into diamond-shaped regions to save computational resources. This algorithm is efficient in obtaining Bayesian PDFs for all prior sources, and thus flux uncertainties can be estimated. However, XID<sup>+</sup> does not use prior information on the FIR flux, and thus might still suffer from some limitations, for example, in the case that there are high-*z* sources that are missing from the prior list but which contribute substantial flux in SPIRE images, or for sources that are close together on the sky but which have very different redshifts (the low redshift ones often will not contribute substantially at SPIRE wavelengths). As an extension to XID<sup>+</sup> photometry, Pearson et al. (2017) developed a method of incorporating galaxy SEDs as “informed priors,” finding improvements in the detection of faint sources. More aggressively, there are also methods that fit multiband images simultaneously by fixing the shape of galaxy SEDs (e.g., MacKenzie et al. 2016) in an approach inspired by Bayesian techniques (Budavári & Szalay 2008; e.g., their Equation (19)). The results from these approaches can be substantially affected by the assumptions about galaxy SED shapes.

From this brief review of earlier photometric methods, it appears that one of the most important steps is choosing the most appropriate prior source list for fitting at each PACS, SPIRE, and (sub)mm band. The prior source list should not be redundant (i.e., it should have about or less than one source per beam; see also Dole et al. 2003; Magnelli et al. 2013). At the same time, it should be as complete as possible, to avoid ignoring any sources that make non-negligible contribution to the total, observed flux.

In this paper we develop a highly optimized method for the extraction of sources in highly confused images, and apply it to *Herschel* PACS and SPIRE and ground-based single-dish (sub)mm images in the GOODS-North field (hereafter

<sup>15</sup> The prior-extraction method adopts a list of prior source positions as the input for the source fitting with PSF or other models to the image data.

<sup>16</sup> The blind-extraction method uses automatic searching algorithms to identify groups of pixels with values significantly greater than the sky background.

<sup>17</sup> Similar to FASTPHOT, or other tools like T-PHOT (Merlin et al. 2015), XID (also named DESPHOT) also uses a linear inversion algorithm as its core fitting function. It additionally adopts a “top-down” approach to iteratively jackknife the actual prior list at each band so as to achieve a best model fit.

<sup>18</sup> Their catalogs are published as Elbaz et al. (2013).

GOODS-N). Our method includes choosing an appropriate prior source list for photometry at each band with the assistance of state-of-the-art SED fitting over all mid-IR to radio bands. Benefiting from the latest understanding of the evolution of IR galaxy SEDs with redshift, and from the availability of high-quality catalogs of optical/near-IR photometric and spectroscopic redshifts, we demonstrate that it is possible to reliably predict the fluxes of galaxies in highly blended FIR bands once flux densities in less blended images have been measured.

These flux density predictions are used exclusively to determine which sources are hopelessly faint and should therefore be excluded (by subtracting their very weak SED-predicted flux densities from the data before further measurements), and which ones are to be kept for the prior-extraction photometry. These flux density predictions do not constrain measurements for the remaining priors; their fluxes are extracted directly from the observed data after the hopelessly faint sources have been flagged and removed. We describe the SED fitting in Section 3.1, and the selection of fitted/excluded prior sources in Section 3.2. The prior-extraction photometry is presented in Section 3.3, and a few examples are illustrated in detail in Section 4.

In addition, we blindly extract sources from the residual image after each photometry step for each band. These additional sources (when detected) have a high likelihood of being high redshift galaxies that are too faint at near-IR and mid-IR (24  $\mu$ m) wavelengths, and thus are not included in our initial prior catalog (although they might also be blends of several low-luminosity lower-redshift sources, or spurious noise peaks). After extracting the additional sources, we re-run the prior-extraction photometry including them. The details of this procedure are given in Section 3.4.

Meanwhile, to verify the performance of the photometry and to provide statistically sound estimates of uncertainties, we run Monte Carlo simulations for each band from 24  $\mu$ m all the way to 1.16 mm and 20 cm. We propose a recipe for correcting formal flux density uncertainties for each individual source for each band, based on a number of directly measurable parameters, calibrated by simulations. The details of the relevant procedures are given in Section 5.

We apply these SED fitting, photometry, simulation, and correction steps to GOODS-N, which is a survey field with some of the deepest and richest multi-wavelength data currently available. The overall flow chart is summarized in Section 6. We present detailed quality checks in Section 7, where we compare our final GOODS-N catalog with several catalogs from the recent literature in Section 7.3, finding that the flux density measurements are generally consistent, while often we have achieved better deblending performance for individual cases. This work also leads to somewhat deeper detection limits in the SPIRE bands. We emphasize, however, that the main advantage of our work is that uncertainties are well-behaved (quasi-Gaussian) and the measurement strategy is, in our opinion, nearly ideal, while still manageable and reproducible, and highly optimized. We believe that the approach described in this paper should be portable to other fields, of course while taking into account the relative depths of the data sets available for prior selection and FIR measurements. A future work (S. Jin et al. 2017, in preparation) will present the application of this technique to the 2-square degree COSMOS field.

We present the final sample of GOODS-N FIR-to-mm (FIR+mm) detected galaxies in Section 8, including photometric measurements, uncertainties, and IR SED photometric redshifts. We derive the dust and star formation properties from the SEDs, and constrain the cosmic star formation rate density (CSFRD) up to redshift 6 in Section 8.6 from directly detected galaxies. Using stellar mass functions (SMFs) from the literature, and assuming the star-forming main-sequence (MS; e.g., Daddi et al. 2007; Elbaz et al. 2007; Noeske et al. 2007; Pannella et al. 2009; Rodighiero et al. 2010; Karim et al. 2011) correlations at each redshift, we estimate completeness correction to the SFR densities. Although the FIR+mm high redshift ( $z > 3$ ) samples have a high incompleteness in stellar mass, they appear to contribute a substantial fraction to the total SFR, and sometimes the number of most massive galaxies at high- $z$  even exceeds what is predicted by empirical SMFs. We further compare rest-frame UV color selections with our sample and discuss how optical to near-infrared studies could miss the dusty galaxies.

Our imaging data sets come from several surveys. MIPS 24  $\mu$ m data come from the GOODS-*Spitzer* program (PI: M. Dickinson). We use coadded PACS images (Magnelli et al. 2013) combining data from PEP (PI: D. Lutz) and GOODS-*Herschel* (PI: D. Elbaz), and SPIRE data from GOODS-*Herschel*. The SCUBA2 data are from the S2CLS program (Geach et al. 2017), and the AzTEC+MAMBO coadded data are from Penner et al. (2011). We used the original SCUBA2 maps rather than the matched-filtered versions (which are also provided by Geach et al. 2017) where the convolution with PSF reduces the effective angular resolution.<sup>19, 20</sup> We used the deep radio imaging from Owen (2017), and the shallower radio imaging from Morrison et al. (2010).

We adopt  $H_0 = 73$ ,  $\Omega_M = 0.27$ ,  $\Lambda_0 = 0.73$ , and a Chabrier IMF (Chabrier 2003) unless specified in the text for specific comparisons to other works. Where necessary, we add  $-0.04$  dex to logarithmic quantities to convert literature measurements from a Kroupa IMF (Kroupa 2002) to a Chabrier IMF, and  $-0.238$  dex to convert from a Salpeter IMF (Salpeter 1955).

## 2. Setting the General Prior Source List

In this section we describe how we set up a prior source list that has enough data points to include, in principle, all possible contributors to the flux densities in FIR to (sub)mm images that we aim to analyze further. As far as we know, there are no distant star-forming galaxies reliably detected in current FIR/submm data that do not also have a counterpart in GOODS-depth IRAC images. This is the natural outcome of the close connection between stellar mass and star formation (i.e., the MS paradigm), coupled with the great sensitivity of *Spitzer* IRAC imaging in GOODS. We therefore start with a list of IRAC-detected galaxies. Most of these galaxies are nevertheless going to be very faint in FIR+mm bands, and we define

<sup>19</sup> As a general comment, there are matched-filtering approaches that do not necessarily imply an increase of the PSF (e.g., by using functions with negative rings like in the method discussed by Chapin et al. 2011, 2013).

<sup>20</sup> We have also carried out our full photometry using the matched-filter version of the SCUBA2 maps, finding a similar detection depth and number of detections as for the unfiltered version. For the matched-filter map, we constructed the PSF image with double-Gaussian function according to the Equation (1) of Geach et al. (2017).



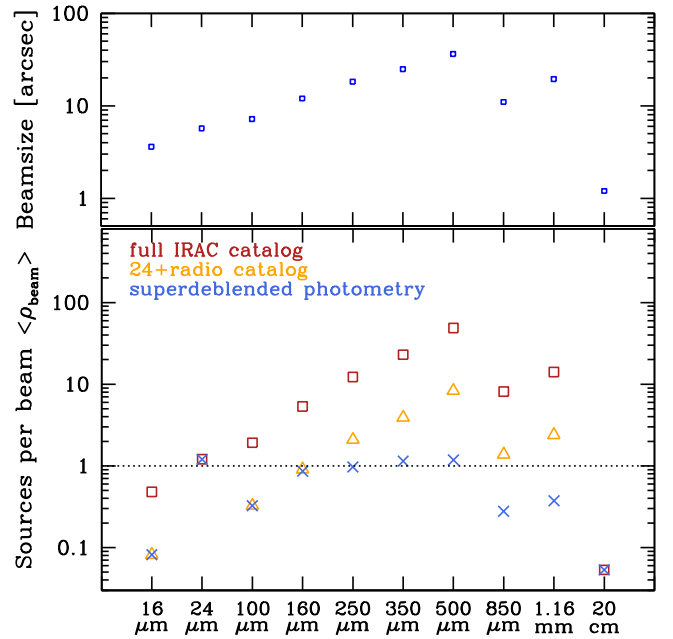
our general prior list as the subset of them that are detected in either 24  $\mu\text{m}$  or 20 cm images, as described below.

We begin with the IRAC source catalog from the GOODS-*Spitzer* project (PI: M. Dickinson), generated using *SExtractor* (Bertin & Arnouts 1996) to detect objects in a coadded IRAC 3.6  $\mu\text{m}$  + 4.5  $\mu\text{m}$  image. The same catalog has also been used as input priors for several previous *Spitzer* and *Herschel* mid-to far-IR catalogs (e.g., Elbaz et al. 2011, 2013; Magnelli et al. 2011, 2013), and contains 19437 IRAC sources. We cross-matched the IRAC source positions with near-IR/optical photometric catalogs using a 1'' separation limit. Primarily, we used the 3D-*HST* catalog (Skelton et al. 2014), which detected sources in *Hubble Space Telescope* (*HST*) WFC3 near-IR data from CANDELS (Grogin et al. 2011; Koekemoer et al. 2011). If no counterpart was found in the 3D-*HST* catalog, we used the catalog of Pannella et al. (2015), which detected sources in CFHT/WIRCAM data from Lin et al. (2012). In this way, we associate near-IR/optical fluxes, photometric redshifts, and stellar masses with each IRAC source. We find overall very good agreement between the two catalogs. A total of 16,862 IRAC sources have photometric redshifts and stellar masses.

We have also cross-matched with spectroscopic redshift catalogs from the literature: Lowenthal et al. (1997), Phillips et al. (1997), Cohen et al. (1996, 2000), Cohen (2001), Steidel et al. (1996, 1999, 2003), Dawson et al. (2001), Barger et al. (2002, 2008), Wirth et al. (2004), Chapman et al. (2004, 2005), Strolger et al. (2004, 2005), Treu et al. (2005), Reddy et al. (2006), Kakazu et al. (2007), Pope et al. (2008), Daddi et al. (2008, 2009a, 2009b, 2010a), Murphy et al. (2009), Hu et al. (2010), Yoshikawa et al. (2010), Cooper et al. (2011), Kajisawa et al. (2011, MODS survey), Stark et al. (2011), Riechers et al. (2011), Casey et al. (2012), Kriek et al. (2015), and Wirth et al. (2015, MOSDEF Survey). Besides, we also use some previously unpublished redshifts from our team (H. Inami et al. 2018, in preparation; D. Stern et al. 2018, in preparation). The references are also numbered in our released catalog. Finally, 16.2% of the IRAC sources have secure spectroscopic redshifts.

In Figure 1, we present the mean number of prior sources per PSF beam area ( $\langle\rho_{\text{beam}}\rangle$ ) for the catalogs used in this work. The IRAC catalog, represented by red squares, is clearly too crowded to be used in its entirety for PSF fitting at PACS and SPIRE bands, resulting in  $\langle\rho_{\text{beam}}\rangle > 10$  at 250  $\mu\text{m}$  and approaching 100 at 500  $\mu\text{m}$ , making deblending virtually impossible. We note in passing that the drop in prior density (which is larger than the drop in beam size) when going from SPIRE to the (sub)mm bands is due to the fact that the 850  $\mu\text{m}$  and 1.16 mm data are effectively shallower than the SPIRE images for sources at the typical redshifts of the priors. However,  $\langle\rho_{\text{beam}}\rangle$  is low enough ( $\lesssim 1$ ) for 16  $\mu\text{m}$ , 24  $\mu\text{m}$ , and 20 cm, so that we can proceed directly to fitting these bands as described below.

We do not remove galaxies classified as passive (e.g., using *UVJ* colors) from the prior pool. In practice, passive sources that might be discarded in this way amount only to a small fraction (of order 10%–15%) of the total. This classification as passive is not always fully reliable, and in some cases galaxies classified as such might actually have positive IR emission (Man et al. 2016; Gobat et al. 2017). By discarding passive galaxy candidates, we could erroneously miss some genuine FIR-emitting galaxies in trade for a fairly small benefit.



**Figure 1.** The upper panel shows the beam sizes (FWHM) of the GOODS-N images in our analysis. The lower panel shows how our prior source selection as described in Section 3 can reduce the number of sources per beam ( $\rho_{\text{beam}}$ ) to about 1 at FIR and mm bands, yielding a balance between source confusion and detection depth. The source densities of the full IRAC catalog are shown as the red squares, the 24  $\mu\text{m}$  and radio-detected sources are orange triangles, while the “super-deblended” prior sources that we selected and actually fitted are shown as blue crosses. We compute the source density by  $\langle\rho_{\text{beam}}\rangle \equiv N_{\text{prior}}/A_{\text{GOODS-N}} \times (\pi/(4\ln 2) \times \theta_{\text{beam}}^2)$ , where  $N_{\text{prior}}$  is the number of prior sources in GOODS-N,  $A_{\text{GOODS-N}}$  is the GOODS-N area ( $10.0 \times 16.5 \text{ arcmin}^2$ ), and  $\theta_{\text{beam}}$  is the beam size (FWHM). The numbers are also listed in Table 1.

## 2.1. PSF Fitting Methodology

In our analysis, we fit source fluxes in the images using *galfit* (Peng et al. 2002, 2010). We use PSF fitting (i.e., we treat the sources as unresolved). The IR data with the highest angular resolution<sup>21</sup> are the *Spitzer* 16 and 24  $\mu\text{m}$  images, with FWHM 4''–6'', a scale at which the intrinsic sizes of most distant galaxies can be neglected. This might not be true for some low redshift galaxies in the images, but these are not the main focus of our efforts.

Since it is not practical to use *galfit* to simultaneously fit a very large number of prior sources in a large image, our code will first divide the image under examination into small regions (boxes). When fitting prior sources in each box, we also consider prior sources from a buffer region of surrounding boxes to avoid edge effects. The buffer size is at least 2–3  $\times$  the PSF FWHM of the image being analyzed. We performed tests to verify that, as long as the box size and buffer size are several times larger than the PSF FWHM, the *galfit* results are unaffected by the exact choices; only the computing speed is affected. We run *galfit* PSF fitting in each box and then combine all boxes to make the full source-model images and residual images. All our PSF fitting is performed at fixed R.A.–decl. positions, as determined from the IRAC catalog, after checking and correcting astrometric differences with each data set under consideration. However, in order to improve the

<sup>21</sup> The PSF FWHM of the VLA data is 2'', and whenever possible we use flux measurements from F. Owen (2018, in preparation) that carefully account for possible extended emission; see the next sections.

fitting for bright sources, we perform a second-pass fit (using first-pass results as first guesses) allowing high S/N (e.g.,  $\gtrsim 10$ ) sources to vary their positions. This second-pass exercise also allows us to evaluate the extra photometric noise (and possible flux biases) introduced by residual (uncorrected) astrometric distortions between IRAC and each IR data set. These terms are generally small but we correct for them in the analysis.

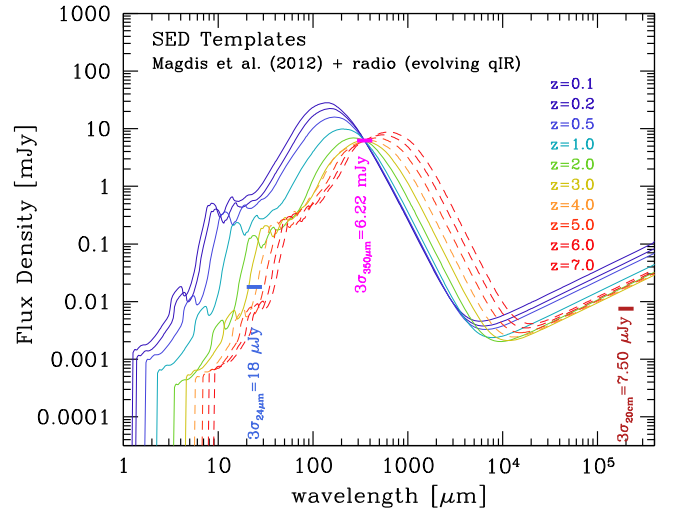
### 2.2. Photometry at 24 $\mu\text{m}$

We fit simple PSFs at 24  $\mu\text{m}$  rather than extended source models because the PSF FWHM at this band is about  $5''.7$ , much larger than the typical sizes of  $z \geq 0.2$  star-forming galaxies, which are  $\sim 8$  kpc or  $2''.5$  at  $z = 0.2$  (e.g., Conselice 2014). For lower-redshift galaxies in the image, we caution that this approach will lead to underestimated flux measurements, for example, for ID 11828, which is the largest spiral galaxy at  $z_{\text{spec}} = 0.106$  in GOODS-N, and for 30 more  $z < 0.2$  FIR-to-mm detected galaxies in our final catalog. To properly remove the background<sup>22</sup> in the 24  $\mu\text{m}$  image, we run a first pass of PSF fitting, then median filter the residual image on a scale of  $30 \times 30$  pixels, then subtract the smoothed background image from the original image. We run a second pass of PSF fitting with fixed prior source positions, and a third pass with varied prior source positions for the highest S/N sources from the second-pass fitting. Then we obtain the final 24  $\mu\text{m}$  fluxes and formal *galfit* errors.

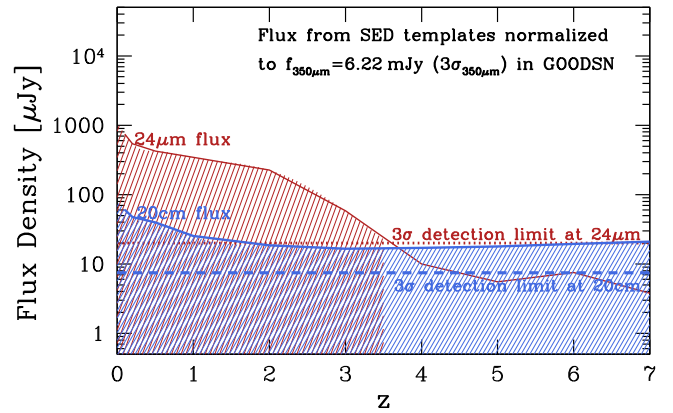
### 2.3. Photometry at VLA 20 cm

The GOODS-N field has very deep VLA  $\sim 20$  cm observations from F. Owen (2018, in preparation) and shallower observations from Morrison et al. (2010).<sup>23</sup> The Morrison et al. (2010) catalog covers a wider region (diameter of  $\sim 15'$ ) than the circular area covered by the F. Owen (2018, in preparation) catalog (diameter of  $\sim 9'$ ). In this work, we use the radio data for two main purposes: to complete the list of prior sources at high redshifts, where galaxies might be too faint to be detected at 24  $\mu\text{m}$ , and to help constrain the overall IR luminosity of each prior source based on the FIR-radio correlation.

The importance of including radio prior sources is illustrated in Figure 2, where we present a series of redshifted SEDs based on templates that we used in this work (see Section 3.1). All SEDs in the figure are normalized to a common  $S_{350 \mu\text{m}} = 6.22$  mJy, which is the  $3\sigma_{350 \mu\text{m}}$  detection limit derived from our simulation (see the following sections). Normalizing to other SPIRE bands will lead to a similar conclusion: the expected  $S_{20 \text{ cm}}$  flux is always somewhat brighter than the empirical detection limit at 20 cm ( $3\sigma_{20 \text{ cm}} \approx 7.5 \mu\text{Jy}$ ), whereas we do not expect to detect the faintest SPIRE sources with  $z \gtrsim 3$  at 24  $\mu\text{m}$ . This is also illustrated in Figure 3, where we plot the predicted flux density at 24  $\mu\text{m}$  and 20 cm as a function of redshift, for the same normalization at 350  $\mu\text{m}$ . The radio data are effectively deeper than the MIPS 24  $\mu\text{m}$  data for selecting sources at  $z \gtrsim 3$ , by roughly a factor of  $\gtrsim 4$ . The predicted radio flux densities are  $\sim 2\times$  brighter than our 20 cm detection limit at any redshift, while the predicted 24  $\mu\text{m}$  flux densities at



**Figure 2.** In order to illustrate prior completeness and depth issues, we show SED templates from the Magdis et al. (2012) library (see Section 3.1; color-coded by their redshifts), which have been normalized to the  $3\sigma$  detection limit at 350  $\mu\text{m}$  that we derive in our analysis:  $S_{350 \mu\text{m}} = 6.22$  mJy (see Table 1). The  $3\sigma$  detection limits at 24  $\mu\text{m}$ , 350  $\mu\text{m}$ , and 20 cm are indicated by short solid bars and vertical text labels. The 24  $\mu\text{m}$  and 20 cm data should be sufficiently deep to provide a complete list of fitting priors for all galaxies with measurable 350  $\mu\text{m}$  fluxes at  $z < 2-3$  (solid line SED templates), largely independently of their detailed SED shapes. At higher redshifts (dashed templates), sources with measurable flux at 350  $\mu\text{m}$  may be undetected at 24  $\mu\text{m}$ , but the 20 cm radio limit is sensitive enough to detect all priors needed to fit the Herschel data. Though valid in general, in practice this may not be true in absolutely all cases, as discussed in Section 3.4.



**Figure 3.** Similar to Figure 2, showing the predicted flux densities at 24  $\mu\text{m}$  (red) and 20 cm (blue) for the Magdis et al. (2012) SED templates as a function of redshift, normalized to the  $3\sigma$  detection limit at 350  $\mu\text{m}$ . The shaded regions indicate the redshift ranges over which 24  $\mu\text{m}$  or 20 cm emission can be detected (i.e., the predicted flux densities are above the detection limits). The flux density at 24  $\mu\text{m}$  fluctuates somewhat with redshift as rest-frame mid-infrared features (e.g., the Polycyclic Aromatic Hydrocarbons; PAHs) move through the passband, and then falls below the detection threshold at  $z > 3.5$ . The flux density at 20 cm slowly decreases with redshift and then flattens to a nearly constant value, and is in principle always detectable for these SPIRE-normalized SED templates that assume the FIR-radio correlation.

$z \gtrsim 4$  are  $\sim 2\times$  fainter than the detection limit. Therefore, by using the VLA 20 cm data, we can have a more complete prior source catalog especially for  $z \gtrsim 3$  sources, improving our completeness for fitting PACS, SPIRE, and (sub)mm photometry.

We use the radio catalog from F. Owen (2018, in preparation) as the main source of 20 cm photometry in this study, with detections down to a  $5\sigma$  level of significance. F. Owen (2018, in

<sup>22</sup> Here and in the rest of the paper we define *background* to be any pedestal level above which sources are emitting, regardless of its origin.

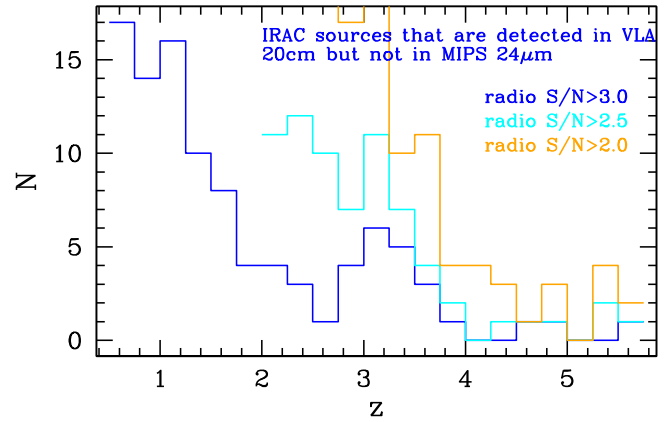
<sup>23</sup> The VLA data from F. Owen (2018, in preparation) were obtained with an average frequency of 1.525 GHz, higher than the 1.4 GHz of Morrison et al. (2010). We convert to 1.4 GHz assuming a canonical radio slope of  $-0.8$  for the comparison and analysis. However, we also note that a radio slope of  $-0.8$  is not applicable to all sources (e.g., Kimball & Ivezić 2008; Marvil et al. 2015).

preparation) carefully derived best estimates for the flux, accounting for possible spatial extension of the sources by comparing fluxes in radio images of different resolution from  $1''$  to  $3''$ . In order to match our prior catalog with IRAC-based positions to the radio we adopt the same matching scheme of F. Owen (2018, in preparation; their Equation (1)), with a tolerance depending on the radio S/N and the sizes of the sources, and including a  $0''.2$  term in quadrature to account for typical IRAC astrometric accuracy.

In order to provide a radio measurement for sources weaker than the F. Owen (2018, in preparation)  $5\sigma$  detection limit (and/or upper limit), we also performed prior-extraction photometry and Monte Carlo simulations independently on the radio imaging data of both Morrison et al. (2010) and F. Owen (2018, in preparation). Small astrometric misalignments (including nonlinear distortions) between the images under exam and the IRAC catalog have been corrected in our photometry by adjusting the prior source positions, at radio, and in all other bands discussed elsewhere in this paper. (Note that this is only important for 24, 16  $\mu\text{m}$ , 20 cm and PACS image data, where the PSFs are still relatively small.) We first run *galfit* to fit PSF models at fixed positions, and then we run *galfit* again, allowing high S/N source positions to vary by a small amount (maximum offset less than 1 pixel). We fit on the radio images without primary beam correction, and then calculate the primary beam correction according to each prior source position and apply it to each photometry measurement.<sup>24</sup>

Note that the PSF of Owen's radio image is about  $\sim 2''$  (provided by F. Owen especially for this work), and  $\sim 1''.7$  for Morrison's radio image.<sup>25</sup> Some sources might be resolved in the radio data, typically at low redshift. In order to verify the performance of our photometry, we compare the measurements derived with our procedure to those from the F. Owen (2018, in preparation) catalog. We find that there is a good agreement in general: the median of the ratio between the different measurements is very close to 1, while the semi-interquartile range is about 6% for bright sources ( $> 80 \mu\text{Jy}$ ), and reaching 10% for fainter fluxes down to the  $5\sigma$  limit. This includes effects of flux losses due to over-resolution of the sources, and is accurate enough for a statistical use of the fluxes from our procedure below the  $5\sigma$  limit, despite its simplicity in assuming unresolved emission and fitting at fixed positions.

As a result of these measurements, we find 1334 IRAC sources with  $S/N_{20\text{ cm}} > 3$  (554 from the Owen 2017 catalog), of which 112 are not detected with S/N of 3 at MIPS 24  $\mu\text{m}$ . The 24  $\mu\text{m}$  undetected radio sources could be radio-loud Active Galactic Nuclei (AGNs), intermediate redshift (e.g.,  $z \sim 1.5$ ) galaxies with strong rest-frame 9.7  $\mu\text{m}$  Silicate absorption (e.g., Magdis et al. 2011), higher redshift star-forming galaxies, although in some cases they could be spurious detections boosted by the noise in the radio data. In Figure 4, we plot the histogram of photometric redshifts for these sources. Indeed two peaks can be found, and the  $z \sim 3$  peak likely represents the high- $z$  star-forming galaxies missed by the MIPS 24  $\mu\text{m}$  selection. We also note that reducing the radio S/N threshold to 2.5 for the high- $z$  (e.g.,  $z > 2$  in the figure) IRAC sources will almost double their number, while making the gap at  $z \sim 2$  shallower (indicated by



**Figure 4.** Histograms of optical/near-infrared photometric redshifts for the IRAC catalog sources that are detected at 20 cm but not at 24  $\mu\text{m}$  (with 24  $\mu\text{m}$   $S/N < 3$ ). These sources could be either radio-loud AGNs, or high redshift ( $z \sim 3$ ) dusty star-forming galaxies (see Section 2.3), or in some cases spurious detections. The blue histogram shows sources with 20 cm  $S/N > 3$ , while the cyan and yellow histograms show the effects of lowering the 20 cm S/N threshold to 2.5 and 2.0, respectively, for sources at  $z > 2$ .

the cyan histogram in Figure 4). Thus in the next section we use 2.5 as the radio S/N threshold for  $z > 2$  sources.

#### 2.4. The 24+Radio Prior Source Catalog

By combining detections at 24  $\mu\text{m}$  and 20 cm, we produce our 24+radio prior source catalog, which will be used for the analysis in the next steps: SED fitting (Section 3.1) and photometry over the FIR-mm bands (Sections 3.2–3.3). There are 2983 sources in the IRAC catalog with  $S/N_{24\text{ } \mu\text{m}} \geq 3$ , plus 112 sources with  $S/N_{24\text{ } \mu\text{m}} < 3$  but with  $S/N_{20\text{ cm}} \geq 3$ .

In order to improve completeness for potential high- $z$  prior sources, which might be detectable in the FIR and (sub)mm data due to their redshifted SED peaks, we also include lower S/N sources (down to  $S/N = 2.5$ ) with photometric redshift above 2.0 in the 24+radio prior source list. In this way, we selected a total of 3306 sources in the 24+radio prior source catalog, with these criteria:  $S/N_{24\text{ } \mu\text{m}} \geq 3$ , or  $S/N_{20\text{ cm}} \geq 3$ , or  $S/N_{24\text{ } \mu\text{m}} \geq 2.5$  when  $z_{\text{phot}} > 2.0$ , or  $S/N_{20\text{ cm}} \geq 2.5$  when  $z_{\text{phot}} > 2.0$ . A total of 47.9% of sources in this full 24+radio prior catalog have spectroscopic redshifts, but only 4.66% of sources have photometric redshifts  $z_{\text{phot}} > 2$ .

Although the flux error measurements at 24  $\mu\text{m}$  and 20 cm are fairly close to Gaussian (see Section 5.3 and the Appendices), starting from a large number of 19,437 IRAC sources would inevitably result in a number of noise-dominated spurious detections at these radio/mid-IR bands, even assuming simple Gaussian statistics. We expect a total of  $\sim 50$  spurious detections above  $3\sigma$  combining the two bands (24  $\mu\text{m}$  and 20 cm). Given that only  $\sim 4000$  IRAC priors have  $z_{\text{phot}} > 2$ , they would also produce some additional  $\sim 50$  sources with spurious radio/mid-IR detection at  $z_{\text{phot}} > 2$  and S/N between 2.5 and 3. These  $\approx 100$  objects would get included among our 3306 priors spuriously, lacking any actual 24  $\mu\text{m}$  or radio detections. This number is small enough that it should not adversely affect the detectability of IR galaxies, as it only marginally increases the number density of priors fitted at all bands. In most cases, we expect these spurious priors to remain undetected in the FIR/mm bands following the super-deblending analysis.

For the remaining 16,131 IRAC sources that were not included in our prior list, we assume that their flux contributions to the PACS, SPIRE, and mm image data are negligible, and we do not

<sup>24</sup> We used the primary beam correction equations on <https://www.cv.nrao.edu/vla/hbg2vla/node41.html>. Most of our correction factors are smaller than 1.5.

<sup>25</sup> The VLA, with its wide bandwidth and multi-frequency synthesis, produces a very clean beam, so the effective beam is very close to a Gaussian.



**Table 1**  
GOODS-N “Super-deblended” Photometry Results

Band	Instrument	Beam FWHM <sup>a</sup> arcsec	$\rho_{\text{fit}}^b$ beam <sup>-1</sup>	$N_{\text{fit}}^c$	$N_{\text{excl.}}^d$	$N_{S/N \geq 3}^e$	$N_{\text{add.}}^f$	$1\sigma^g$ mJy
24 $\mu\text{m}$	<i>Spitzer</i> /MIPS	5.7	1.205	19437	0	3056	0	$5.165 \times 10^{-3}$
20 cm	VLA	1.7/2.0 <sup>h</sup>	0.107	19437	0	1328	0	$2.744 \times 10^{-3}$
16 $\mu\text{m}$	<i>Spitzer</i> /IRS/PUI	3.6	0.082	3306	0	1335	0	$7.681 \times 10^{-3}$
100 $\mu\text{m}$	<i>Herschel</i> /PACS	7.2	0.326	3294	12	1178	0	0.315
160 $\mu\text{m}$	<i>Herschel</i> /PACS	12.0	0.862	3137	169	1153	18	0.681
250 $\mu\text{m}$	<i>Herschel</i> /SPIRE	18.2	0.973	1540	1766	668	13	1.571
350 $\mu\text{m}$	<i>Herschel</i> /SPIRE	24.9	1.142	966	2340	292	10	2.072
500 $\mu\text{m}$	<i>Herschel</i> /SPIRE	36.3	1.181	470	2836	125	17	2.570
850 $\mu\text{m}$	JCMT/SCUBA2	11.0 <sup>i</sup>	0.154	668	2638	32	15	1.249
1160 $\mu\text{m}$	AzTEC+MAMBO	19.5	0.374	515	2791	43	11	0.661

**Notes.**

<sup>a</sup> Beam FWHM is the full width at half maximum of the circular-Gaussian-approximation point-spread function of each image data.

<sup>b</sup>  $\rho_{\text{fit}}$  is the number density of prior sources fitted at each band, normalized by the Gaussian-approximation beam area.

<sup>c</sup>  $N_{\text{fit}}$  is the number of prior sources fitted at each band.

<sup>d</sup>  $N_{\text{excl.}}$  is the number of prior sources excluded from fitting at each band. These sources are subtracted from original image with their SED-predicted flux density at each band.

<sup>e</sup>  $N_{S/N \geq 3}$  is the number of prior sources with  $S/N \geq 3$  (i.e., detected) at each band.

<sup>f</sup>  $N_{\text{add.}}$  is the number of  $S/N \geq 3$  additional sources that are not in the prior source catalog but blindly extracted from the intermediate residual image product at each band (see Section 3.4).

<sup>g</sup>  $1\sigma$  is the median of the flux density uncertainties of all sources detected with  $S/N \geq 3$  in each band.

<sup>h</sup> The shallower VLA image data from Morrison et al. (2010) have a beam FWHM of 1".7, and the deeper VLA image data from F. Owen (2018, in preparation) are especially produced with a beam FWHM of 2".0.

<sup>i</sup> We use the unfiltered image, which has a narrower PSF FWHM than that of the matched-filter convolved image. We measured  $\text{FWHM} \approx 11''$  in our PSF reconstruction for the unfiltered image, which is roughly consistent with Figure 14 of Chapin et al. (2013), and compares with 14".8 for the matched-filter image.

consider them further in the rest of this work. Even though their fluxes will not be exactly zero, their presence will act as a background whose average level will be consistently taken into account by our procedure. Their possibly inhomogeneous distribution will act as additional confusion noise, which our procedure will also quantify.

Figures 2 and 3 suggest that our prior source list is fairly complete and will include, in principle, any FIR-mm detectable star-forming galaxy, even at the highest redshifts, thanks to our very deep 20 cm detection limit. This of course is no longer entirely true when scatter in the FIR-radio correlation and SED shape variations are considered, instead of assuming galaxies to be perfectly represented by Magdis et al. (2012) templates, as in Figures 2 and 3. Also, the presence of noise in the measurements might affect the detectability of sources close to the detection limits. Therefore, despite using the deepest 24  $\mu\text{m}$  and radio catalogs, we do expect some incompleteness in our prior list for  $z > 3$  galaxies at the faintest (but detectable) flux levels, especially over the SPIRE bands and in the (sub)mm. This motivates a posteriori addition of sources found in residual map (mainly) in these bands, as will be discussed in Section 3.4.

The combination of 24  $\mu\text{m}$  and radio priors is also useful to overcome the limitations at 24  $\mu\text{m}$  where we fit all IRAC priors with a density of about 1 source per beam. This will imply that the effective depth of our catalog at 24  $\mu\text{m}$  will depend on local crowding (see further discussions about *crowdedness* in this paper). However, the beam at 20 cm is small enough that crowding is irrelevant. Using the radio data should reduce the possibility that we are missing useful priors in regions where 24  $\mu\text{m}$  is most crowded.

A more sophisticated way to complete the prior sample at high redshifts would be to cull more sources from the IRAC catalog and to use the correlation between stellar mass and SFR to single

out appropriate additional high redshift priors. We have carried out a preliminary investigation of this idea in GOODS-N, and find that this does not provide an obvious improvement to what has been achieved in the current work. The stellar mass selection will be explored in more detail in future work in COSMOS (S. Jin et al. 2018, in preparation), where the prior catalogs at 24  $\mu\text{m}$  and radio wavelengths are shallower than in GOODS-N.

### 2.5. Photometry at 16 $\mu\text{m}$

Despite having a fairly small PSF ( $\text{FWHM} \approx 4''.5$ ; hence a small number of IRAC sources per beam), the GOODS-N 16  $\mu\text{m}$  imaging data observed by the *Spitzer* IRS Peak-Up Imaging (PUI) are a factor of  $\simeq 2$  shallower (for the same SFR) than the MIPS 24  $\mu\text{m}$  image data ( $\sim 7.5 \mu\text{Jy}$  versus  $\sim 5 \mu\text{Jy}$  sensitivities; see Table 1). Therefore in this work we directly fit the 16  $\mu\text{m}$  image data at all 24+radio prior positions to measure 16  $\mu\text{m}$  fluxes. Only 40% of our 3306 24+radio prior sources have  $S/N_{16 \mu\text{m}} \geq 3$ .

## 3. Super-deblended Photometry for Blended FIR/mm Images

In this section we describe the core of our photometric method to obtain super-deblended photometry in confused FIR/mm images.

We proceed one band at a time, working toward longer wavelengths and (generally) larger beam sizes. For example, at this stage, we have obtained 16, 24  $\mu\text{m}$  and 20 cm photometry for the 3306 24+radio sources in our full prior list derived from a parent IRAC catalog. We use this information to optimize photometry in the next band, PACS 100  $\mu\text{m}$ , and continue similarly for other bands.



We use SED fitting of all available photometry at each stage to predict the flux at the next FIR/mm band, with the aim of optimizing the pool of prior sources to be fitted at each FIR/mm band. Once we have predicted fluxes and uncertainties for all sources in a given band, we determine a criterion to distinguish faint sources (whose fluxes are then subtracted from the original images), from those that are retained and eventually used as prior positions for flux measurement by PSF fitting.

We note that in this way, the “earlier” bands are simply used to decide which priors to use for *galfit* analysis of the data. PSF fitting with *galfit* is carried out one band at a time (i.e., we are not making photometric measurements simultaneously in multiple bands). Such an approach of simultaneous multiband fitting could have benefits, but also has disadvantages, notably the fact that complex inter-band dependencies would be added by the need to use some kind of SEDs to connect information across bands, making it much harder to reliably derive photometric uncertainties for each band.

We emphasize that if a source is discarded for fitting at any given band (e.g., say at 100  $\mu\text{m}$  because it is predicted to be too faint there), this does not exclude the possibility that the same source might be fitted in further steps at longer wavelengths (e.g., in SPIRE or SCUBA2). Indeed, this is quite likely to happen for prior sources with a high redshift.

Finally, we inspect the residual images to perform blind detection and extraction of sources that might still be present after all sources fit with priors have been removed. In later sections, we provide a detailed example of the procedure for galaxies in the area of GN20 (Daddi et al. 2009b), and some comparison of our measurements to published catalogs.

We have applied a factor of 1.12 $\times$  to the final PACS flux densities and uncertainties (used in this paper and released in our catalog) to account for the flux losses from the high-pass filtering processing of PACS images (e.g., Popesso et al. 2012; Magnelli et al. 2013).

### 3.1. SED Fitting Algorithm

We consider four distinct SED components in fitting procedure. From shorter to longer wavelengths, these are: (1) a stellar component from Bruzual and Charlot (2003, hereafter BC03) with a Small Magellanic Cloud (SMC) attenuation law; (2) a mid-infrared AGN torus component from Mullaney et al. (2011); (3) dust continuum emission as modeled by the Magdis et al. (2012) library; and (4) a power-law radio component (see the text below).

The Magdis et al. (2012) library is based on Draine and Li (2007, hereafter DL07) templates fitted to the average SEDs of MS and starburst (SB) galaxies as a function of redshift. DL07 templates are parametrized by a number of physical properties: the minimum interstellar radiation field (ISRF) intensity  $U_{\text{min}}$ , the maximum ISRF intensity  $U_{\text{max}}$ , the dust mass  $M_{\text{dust}}$ , the mass fraction of PAH to total dust mass  $q_{\text{PAH}}$ , and the mass fraction of dust grains located in photo-dissociation regions (PDR)  $\gamma$ , and so on. Hence they provide quite a wide range of SED shapes. However, Magdis et al. (2012) simplified the parametrization of DL07 templates for galaxies at various redshifts with only two parameters: the IR luminosity per dust mass  $L_{\text{IR}}/M_{\text{dust}}$  or the mean ISRF intensity  $\langle U \rangle$ ,<sup>26</sup> and whether the source is on the MS or not. Magdis et al. (2012) find that the shapes of the dust SEDs of MS galaxies at a given redshift, as traced by  $L_{\text{IR}}/M_{\text{dust}}$ , is not

expected to vary significantly with increasing  $L_{\text{IR}}$  or SFR. Meanwhile,  $L_{\text{IR}}/M_{\text{dust}}$  changes as a function of redshift, and the variation of SED shape among MS galaxies is expected to depend only on  $L_{\text{IR}}/M_{\text{dust}}$ , or equivalently  $\langle U \rangle$ . Magdis et al. (2012) constructed a series of dust SED templates for MS galaxies parametrized only by  $\langle U \rangle$ , as shown in their Figure 16. These templates assumed an evolution of  $\langle U \rangle \propto (1+z)^{1.15}$ , no evolution in PDR fraction, and a small evolution of  $q_{\text{PAH}}$  beyond  $z = 2$ , as indicated by their data. Béthermin et al. (2015) updated the evolution of  $\langle U \rangle \propto (1+z)^{1.8}$  based on the latest data in the COSMOS field. Therefore, in this work, we use the Magdis et al. (2012) dust SED templates (which depend on  $\langle U \rangle$ ) with redshift evolution from Béthermin et al. (2015) (which determines  $\langle U \rangle$  at each redshift) to fit galaxy SEDs and predict photometric redshift and FIR/mm fluxes. Examples of our SED templates are shown in Figure 2.

Meanwhile, at all redshifts, there is a small fraction of SB galaxies which have SFRs higher than those of MS galaxies with similar stellar masses. Their  $L_{\text{IR}}/M_{\text{dust}}$  ratios or  $\langle U \rangle$  are found to be higher than those for MS galaxies (at least at  $z < 1.5$ ) and likely do not vary with redshift (Magdis et al. 2012; Tan et al. 2014; Béthermin et al. 2015). So their SEDs can be also parametrized in the same way, by a constant  $\langle U \rangle$ . When fitting MS galaxies at a given redshift, we allow for a range of  $\langle U \rangle$  values spanning  $\pm 0.2$  dex around the expected value, to allow for the observed scatter of dust temperature among MS galaxies (defined by  $z = 0$  results; see Magdis et al. 2012).

A power-law radio SED has been added to describe the radio continuum as  $(S_\nu \propto \nu^{-0.8})$ <sup>27</sup> and tied to  $L_{\text{IR}}$  by the FIR-radio logarithmic ratio  $q_{\text{IR}} \equiv \log_{10}(S_{\text{IR}}, 8\text{--}1000 \mu\text{m}/3.75 \times 10^{12} \text{ W m}^{-2}) - \log_{10}(S_{1.4 \text{ GHz}}/\text{W m}^{-2} \text{ Hz}^{-1})$  (e.g., Helou et al. 1985; Condon 1992; Ivison et al. 2010; Sargent et al. 2010; Tan et al. 2014), so that the radio data point can be used in the fit together with FIR/mm to constrain  $L_{\text{IR}}$ . Here we adopt the latest results that suggest a slowly evolving  $q_{\text{IR}} = 2.35 \times (1+z)^{-0.12} + \log(1.91)$ , following Magnelli et al. (2015) and Delhaize et al. (2017).<sup>28</sup> We find that such a small evolution is also generally warranted by our global SEDs. We verified that adopting a constant  $q_{\text{IR}}$  would have negligible impact on the photometric analysis carried out in this paper.

#### 3.1.1. SED Fitting Parameters

In the case of radio-excess sources or radio-loud AGNs, which are outliers of the FIR-radio correlation, the radio emission does not predominantly arise from star formation. In these cases the radio photometry should not be used in SED fitting. Also, we find that mid-IR rest-frame emission is at times hard to fully reproduce with models that fit the FIR emission. For this and other analogous situations in the following, we introduce three parameters which summarize the SED fitting approach and which are included in our final catalog: (1) a flag to distinguish between MS and SB galaxies; (2) a flag to identify radio-excess; and (3) a flag for high-quality FIR photometry (in

<sup>26</sup>  $\langle U \rangle$ , by its definition, is proportional to  $L_{\text{IR}}/M_{\text{dust}}$  (see, e.g., Magdis et al. 2012).

<sup>27</sup> We choose a slope of  $-0.8$  in this work (e.g., Kellermann & Owen 1988; Kimball & Ivezić 2008). But we also note that a single slope may not apply to all galaxies (e.g., Marvil et al. 2015). Since we use radio data mainly for optimizing the fitted prior sources at FIR/mm band, and then focus our attention on the FIR/mm properties of the sources, we do not implement a galaxy-dependent radio slope for the SED fitting.

<sup>28</sup>  $\log(1.91)$  accounts for the conversion between  $L_{\text{FIR}}$  (rest-frame 42–122  $\mu\text{m}$ ) and  $L_{\text{IR}}$  (rest-frame 8–1000  $\mu\text{m}$ ) according to Magnelli et al. (2015).

which case radio and 24  $\mu\text{m}$  fluxes are not used in the SED fitting and flux prediction; they are unnecessary and can possibly be misleading if they are affected by a radio-loud or mid-IR AGN).

The first parameter is the MS/SB classification (noted as “Type\_SED”), corresponding to the use of MS/SB types of SED templates. This is determined by carrying out the SED fitting twice, as we need a first estimate of SFR to decide if the galaxy is MS or SB. In the first pass, we fit all MS and SB SED templates for each source and derive the best fitting values and uncertainties for  $z_{\text{phot}}$ ,  $\langle U \rangle$ ,  $M_{\text{dust}}$ , and  $L_{\text{IR}}$ , based on the equations in Press et al. (1992). SFRs are computed from the integrated  $L_{\text{IR}}$ , assuming a Chabrier IMF (Chabrier 2003) and the relation  $\text{SFR} = L_{\text{IR}} / (1 \times 10^{10} L_{\odot}) M_{\odot} \text{ yr}^{-1}$  (Daddi et al. 2010b). We then compare the SFR and its uncertainty  $\sigma_{\text{SFR}}$  with the MS-based SFR ( $\text{SFR}_{\text{MS}}$ ) according to Equation (A1) of Sargent et al. (2014):  $\text{SFR}_{\text{MS}} = M_{*} \times N(M_{*}) \times \exp\left(\frac{A \cdot z}{1 + B \cdot z^C}\right)$ , where the redshift  $z$  comes from the SED fitting and the stellar mass  $M_{*}$  is from optical/near-IR catalogs, as described in Section 2 when available. The parameter values  $N_{M_{*}}$ ,  $A$ ,  $B$ , and  $C$  are taken from Sargent et al. (2014, appendix) and are described therein. Sources are set to pure SB (Type\_SED = 1) if  $\log(\text{SFR}/\text{SFR}_{\text{MS}}) > 0.6$  dex and  $\text{SFR}/\sigma_{\text{SFR}} > 3$ , and to pure MS (Type\_SED = -1) if  $\log(\text{SFR}/\text{SFR}_{\text{MS}}) < 0.4$  dex and  $\text{SFR}/\sigma_{\text{SFR}} > 3$ . For all other sources (e.g., those without  $M_{*}$  estimate or those for which the SFR does not meet the criteria above), we assume that we cannot conclusively decide whether they are MS or SB galaxies, and we set Type\_SED = 0, using all MS and SB templates to search for the best fit to their SEDs.

The second parameter is the classification of radio-excess sources or radio-loud AGNs (noted as “Type\_20 cm”). This is also determined using a two-pass SED fitting approach. In the first pass, we do not use the observed radio flux density ( $S_{\text{OBS,radio}}$ ). Since the radio SED component is tied to the  $L_{\text{IR}}$  by the FIR-radio correlation, we can derive an expected radio flux density ( $S_{\text{SED,radio}}$ ) and associated uncertainty from the IR SED fit. Then, the (dis)agreement between  $S_{\text{OBS,radio}}$  and  $S_{\text{SED,radio}}$  determines our choice of the parameter Type\_20 cm:

if  $(S_{\text{OBS,radio}} - S_{\text{SED,radio}}) / \sqrt{\sigma_{\text{OBS,radio}}^2 + \sigma_{\text{SED,radio}}^2} > 3$  and  $S_{\text{OBS,radio}} > 2 \times S_{\text{SED,radio}}$ , then we set Type\_20 cm = 1; otherwise Type\_20 cm = 0. Then, in the second pass, if a source has Type\_20 cm = 1, we do not use its radio data, while for the remaining sources we use the radio data to provide tighter constraints on the physical properties. Overall, we have flagged 10.3% of the 24+radio source as radio-excess/radio-loud candidates. Our approach is very conservative, flagging a source as a possible radio AGN when both the bolometric  $L_{\text{IR}}$  and radio are well measured and when the radio flux exceeds the value predicted by the bolometric  $L_{\text{IR}}$  by a factor of two or more, which is comparable to the scatter of the radio-IR correlation. Some of these sources will not actually be AGNs, but we refrain from using their radio fluxes for SED fitting when we have a well-determined IR SED.

The third parameter is the combined S/N computed over the FIR/mm bands (noted as “Type\_FIR”). For sources with FIR/mm combined S/N  $\geq 5$ ,<sup>29</sup> we set Type\_FIR = 1 and do not fit the 24  $\mu\text{m}$  and radio data points anymore. In this way our SED fitting procedure will not be affected by the scatter of the

FIR-radio relation, or by galaxy–galaxy variations of the mid-IR dust SED features (e.g., variations of the IR to rest-frame 8  $\mu\text{m}$  ratio, IR8, Elbaz et al. 2011). This optimization improves the FIR/mm part SED fitting of individual sources, but does not lead to obvious difference in the statistical results in Section 8.

As has already been briefly mentioned, the SED fitting is (re) run for each FIR/mm band prior to extracting photometry for that band. For example, before we measure PACS 100  $\mu\text{m}$  fluxes, we run SED fitting to the  $K_s$ , IRAC, 16  $\mu\text{m}$ , 24  $\mu\text{m}$ , and 20 cm (if not radio-excess/radio-loud) data points for each 24+radio prior source. Similarly, before extracting SPIRE 350  $\mu\text{m}$  photometry, we run SED fitting with the aforementioned bands plus the newly measured 100, 160, and 250  $\mu\text{m}$  data points. The purpose of this SED fitting is to provide the best possible prediction for choosing the final prior sources that are used for fitting. This is a crucial step toward obtaining super-deblended photometry.

For sources that have reliable spectroscopic redshifts, we fix the redshifts to those values for the SED fitting. For those with optical/near-IR photometric redshifts, we perform SED fitting, allowing the redshift values to span a range of  $\pm 10\%$  in  $(1 + z)$ , corresponding to an uncertainty range of about  $\pm 2\sigma$  (Skelton et al. 2014; Pannella et al. 2015). We do not allow redshifts to vary beyond this range, even if the minimum  $\chi^2$  were to be found at the edge of the interval, to avoid getting solutions that are too inconsistent with the optical constraints. For IRAC sources without any available optical/near-IR photometric redshift ( $\sim 10\%$  of the total IRAC catalog, but only 2.3% of the 3306 24+radio priors<sup>30</sup>), we allow the full range  $0 < z < 8$  for the SED fitting. The IR-to-radio SED redshifts are presented as  $z_{\text{IR}}$  in our catalog (e.g., Table 2).

For each band, the SED fitting procedure returns a most likely expected flux density ( $S_{\text{SED}}$ ), as well as its uncertainty ( $\sigma_{\text{SED}}$ ) based on the  $\chi^2$  statistics. This procedure requires that we can associate a well-behaved, (quasi-)Gaussian uncertainty to the fluxes measured in all shorter-wavelength bands. The description of the derivation of reliable flux uncertainties will be presented in Section 5.

### 3.2. Optimized Prior Source Lists for Each FIR/mm Band

Figure 5 shows the number of prior sources per beam with  $S_{\text{SED}} + 2\sigma_{\text{SED}}$  above threshold values indicated by the X axis. We add twice the flux density uncertainty to the best-fit flux density in order to conservatively estimate a plausibly maximum flux density for each source. For each band we show a black vertical line that indicates the median  $1\sigma$  flux density uncertainty that we finally obtain from our procedure (see Table 1).

Using these figures, we can optimally determine the threshold value  $S_{\text{cut}}$  for our analysis for each band. We recall that sources with  $S_{\text{SED}} + 2\sigma_{\text{SED}}$  larger than  $S_{\text{cut}}$  are included in the list of prior sources for *galfit* photometric measurements in a given band (henceforth, *selected* sources, for that band). Sources with  $S_{\text{SED}} + 2\sigma_{\text{SED}}$  fainter than  $S_{\text{cut}}$  in a given band are excluded from *galfit* fitting in that band (henceforth, *excluded* sources), and their flux densities, as predicted from the SED fitting, will be subtracted from the image, which will be discussed in the next section. A larger value of  $S_{\text{cut}}$  will lead to less confusion in *galfit* fitting (i.e., a smaller number density of

<sup>29</sup> For example, at 350  $\mu\text{m}$  we calculate the combined S/N from 100  $\mu\text{m}$  to 250  $\mu\text{m}$  as  $\sqrt{S/N_{100\mu\text{m}}^2 + S/N_{160\mu\text{m}}^2 + S/N_{250\mu\text{m}}^2}$ , and similarly at other bands.

<sup>30</sup> These could be real extremely red galaxies, or perhaps blends with displaced centroids that could affect cross-matching to the optical data. Further exploration of this is beyond the scope of this paper.

**Table 2**  
GOODS-N “Super-deblended” Photometry Catalog (Example)

ID (1)	R.A. (2)	Decl. (3)	$z_{\text{IR}}$ (4)	$z_{\text{type}}$ (5)	$S_{500}$ (6)	$\sigma_{500}$ (7)	$S/N_{\text{IR}}$ (8)	$\log M_*$ (9)	$\text{SFR}_{\text{total}}$ (10)	$\text{SFR}_{\text{IR}}$ (11)	$\sigma_{\text{SFR}_{\text{IR}}}$ (12)	goodArea (13)	$T_{20 \text{ cm}}$ (14)	$T_{\text{SED}}$ (15)
564	189.2994995	62.3700371	4.055	s	44.0	5.8	19.9	11.25	1812.4	1726.0	86.6	1	0	1
11499	189.2606049	62.2172966	4.50	p	10.4	1.8	18.2	11.44	983.5	977.4	53.7	1	0	-1
2592	189.3670044	62.3222733	4.74	p	14.5	2.3	16.3	11.37	1238.9	1231.0	75.4	1	0	-1
15289	188.9900818	62.1734276	3.075	s	8.7	2.3	15.4	8.06	386.1	354.9	25.5	0	0	1
16332	189.2442017	62.1585388	3.55	p	11.3	3.8	14.9	...	659.8	594.8	163.5	1	0	0
3532	189.3077393	62.3073006	3.99	p	9.1	2.2	14.8	11.46	689.1	687.9	190.2	1	1	-1
9053	189.0358124	62.2431564	3.48	p	14.7	1.8	14.4	11.26	579.2	576.1	88.5	1	0	-1
4500	189.4094086	62.2934532	3.190	s	14.6	2.5	14.1	10.26	450.7	317.7	22.5	1	0	1
18911	189.1133881	62.1015778	4.55	p	31.8	3.9	14.1	...	1323.3	1193.0	100.4	0	0	0
4990	189.1330261	62.2873993	4.16	p	12.6	2.3	13.8	10.32	698.5	687.5	49.8	1	0	1
14914	188.9597931	62.1782951	5.33	p	20.1	2.6	13.6	12.55	1760.0	1724.0	132.1	0	0	-1
15213	189.0117798	62.1743240	3.24	p	5.9	2.9	12.3	10.45	399.7	384.2	148.9	1	0	0
17624	189.0368347	62.1344681	3.36	p	12.1	4.0	11.7	11.46	512.1	511.2	88.0	0	0	-1
16810	189.0555725	62.1504669	3.05	p	10.3	1.6	11.4	10.84	338.3	327.8	71.1	0	0	-1
13107	189.1156006	62.1996002	3.49	p	9.2	1.9	11.3	11.07	433.2	427.9	128.4	1	0	-1
130	189.3960114	62.3908005	4.13	p	23.7	2.5	11.0	10.67	1109.8	1085.0	98.5	0	0	1
3827	189.5537872	62.3028259	3.90	p	29.0	3.4	10.8	...	1083.8	977.1	264.0	0	0	0
16121	189.1436768	62.1616745	3.74	p	12.2	2.1	10.6	11.24	454.0	446.2	42.1	1	0	-1
17381	189.2108612	62.1394119	3.10	p	3.8	9.3	10.5	10.68	491.3	445.9	160.4	1	0	0
18603	189.0650787	62.1119728	3.72	p	12.6	6.6	10.4	11.04	742.8	737.7	163.4	0	0	0

**Note.** Here we show a few example columns and a few example  $S/N_{\text{FIR+mm}} \geq 10$  sources at  $z \geq 3$  (sorted by the  $S/N_{\text{FIR+mm}}$ ). Column (1)–(3) are the IRAC catalog from GOODS-*Spitzer* Legacy Program (see Section 2). Column (4) is our IR-to-radio SED redshift, which is spectroscopic redshift if available and otherwise the SED fitted photometric redshift; see Section 3.1. Column (5) indicates whether  $z_{\text{IR}}$  is spectroscopic redshift (“s”) or photometric redshift (“p”). The references for each spectroscopic redshifts are fully listed in the released machine-readable catalog. Column (6) and (7) are flux density and uncertainty in unit of mJy. Column (8) is the FIR+mm combined  $S/N$  in Equation (2). Column (9) is the stellar mass from 3D-*HST* Skelton et al. (2014) or Pannella et al. (2015), converted to Chabrier IMF when needed. Column (10) is the total SFR, adding UV-unattenuated SFR (Section 8.2) to the dust-obscured SFR. Column (11) and (12) are the dust-obscured SFR and uncertainty from FIR+mm SED fitting (see the Figure 23 caption for deriving SFR from  $L_{\text{IR}}$ ). Column (13) is the “goodArea” parameter, equals 1 if the source is in the inner lower rms area in 24  $\mu\text{m}$  (has deeper prior catalog for surrounding sources), and otherwise is 0 (has shallower or even incomplete prior catalog for surrounding sources). Column (14) is the “Type\_20 cm” parameter, equals 1 if the source is classified as radio excess, and its radio data point is not fitted in SED fitting, and otherwise is 0. Column (15) is the “Type\_SED” parameter, equals 1 if the source is classified as starburst type (its SED is fixed to using starburst templates), or -1 if classified as MS type (its SED is fixed to using MS templates), and otherwise is 0 (allowing fitting all SED templates).

(This table is available in its entirety in machine-readable form.)

fitted sources,  $\rho_{\text{beam}}$ ) but a larger number of fainter sources to be subtracted, and vice versa. Also, setting too large a value for  $S_{\text{cut}}$  might prevent us from obtaining measurements for galaxies with detectable FIR/mm emission. We therefore set  $S_{\text{cut}}$  by requiring both  $\rho_{\text{beam}} \lesssim 1$  and also satisfying  $1\sigma \lesssim S_{\text{cut}} \lesssim 3\sigma$  at each band, where  $\sigma$  is the empirical median noise per beam at this band calculated from simulations as described below. The lower boundary is set to  $1\sigma$  to avoid fitting sources that are too faint for the data in hand. The upper boundary of  $3\sigma$  is still faint enough (particularly considering that we have added  $2\sigma_{\text{SED}}$  to the SED flux densities when evaluating the cutoff threshold) that we do not run much risk of discarding priors for any sources that are individually detectable in the data.

Regardless of these criteria, we also always keep as a prior any source that was significantly detected in the previous (shorter-wavelength) band with  $S/N > 3$ . In this way, for sources that have high  $S/N$  in the preceding band but that are predicted to be too faint in band currently under analysis, we can still obtain an upper limit that is useful in SED fitting.

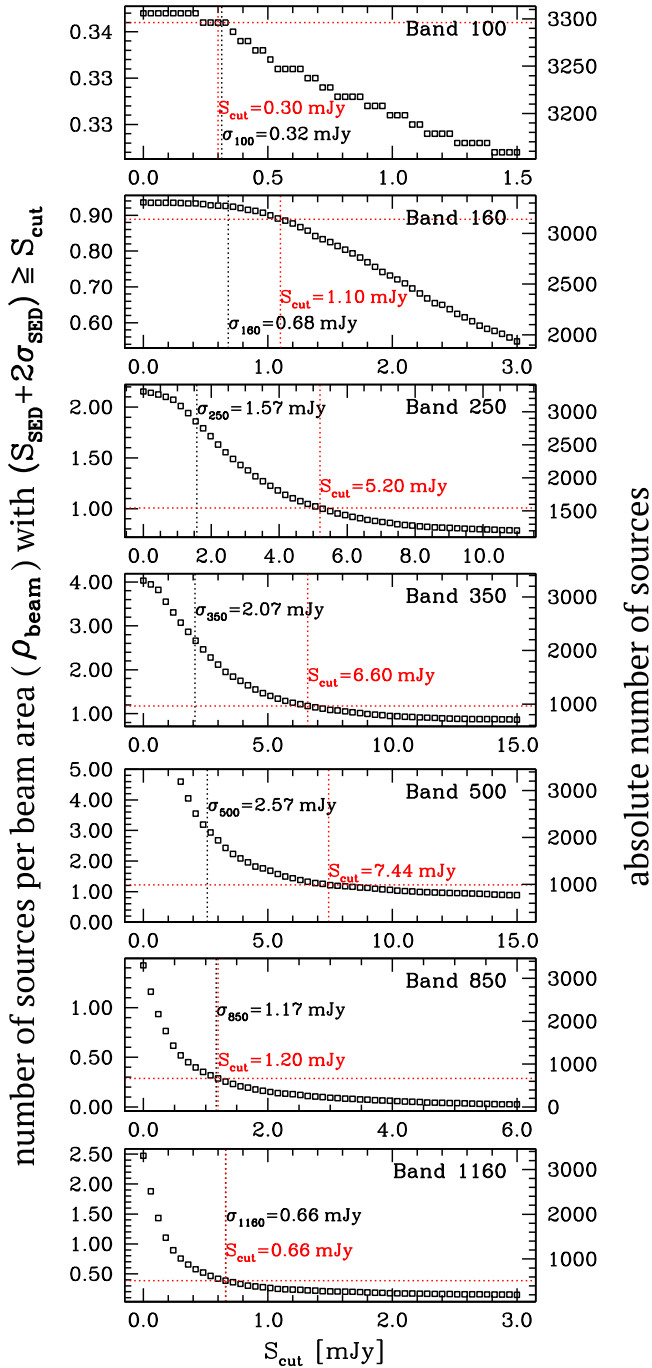
Figure 5 shows the effect of imposing  $S_{\text{cut}}$  thresholds on each band. The number of the *selected* sources is reduced to only  $\sim 30\%$  of the total number (3306) at 350  $\mu\text{m}$  and 500  $\mu\text{m}$  bands (overcoming blending), and to  $\sim 15\%$  at 1.16 mm (which avoids fitting overly faint sources).

### 3.3. Subtract Excluded Sources and Fit Selected Sources

Sources that are excluded from *galfit* PSF fitting through the previous section analysis do still contribute some weak flux in the observed images, and could thus boost the photometry for sources that are retained for fitting. Therefore we first make a model image of those excluded sources by fixing their fluxes to the predicted values, and then subtract the faint-source-model image from the observed original image and obtain a faint-source-subtracted image. The *galfit* photometry is then performed on the faint-source-subtracted image for the selected prior sources in a similar way, as described in Section 2.2. In Figure 6, we present the SPIRE 250  $\mu\text{m}$  images in each stage of this procedure. In the Appendix we show all of the images produced during these steps for all of the bands analyzed in this paper. It is interesting to compare the total flux subtracted from the original images (for faint excluded sources) with that actually extracted from the fitted sources. This reaches up to 30% of the total flux in the worst cases for the longer-wavelength SPIRE bands, where only 15%–20% of the original priors are actually fitted, and is generally much smaller for other bands.

The list of prior sources that are selected for fitting extends to sufficiently faint (i.e.,  $< 3\sigma$ ) flux densities for each band that we do not expect all sources used for the PSF fitting to actually be





**Figure 5.** Cumulative number density  $\rho_{\text{beam}}$  of prior sources vs. the threshold (cut) value applied to their expected flux densities, conservatively increased by twice the flux density uncertainty, as discussed in the text. This is conservative because the large majority of galaxies will have actual flux densities lower than this value. We select sources for PSF fitting with *galfit* PSF when their flux densities are larger than the cut value:  $S_{\text{SED}} + 2\sigma_{\text{SED}} \geq S_{\text{cut}}$ , where  $S_{\text{SED}}$  is the SED best fitting flux density and  $\sigma_{\text{SED}}$  is the SED flux density uncertainty derived from  $\chi^2$  distribution statistics (Section 3.1). The left-Y axis  $\rho_{\text{beam}}$  has the same definition as in Figure 1. The right-Y axis indicates the total number of sources for fitting at each band. The posterior  $1\sigma$  detection limit for each band is labeled with black text in each panel. The flux density cut value  $S_{\text{cut}}$  is labeled in red.

detected with  $S/N > 3$  in the extraction process. The number of  $S/N > 3$  detections ( $N_{S/N>3}$ ) is generally of order of 1/3 of all fitted sources ( $N_{\text{fit}}$ ).  $N_{\text{fit}}$  and  $N_{S/N>3}$  of all bands are listed in Table 1.

### 3.4. Additional Sources in the Residual Image

Although the 24+radio combination provides a very deep prior source catalog for fitting the FIR/mm bands, some less massive, radio-quiet high- $z$  galaxies could still be missed. For example, in Figures 2 and 3, the  $z \sim 3$ –5 SEDs have 24  $\mu\text{m}$  and 20 cm fluxes that are close to or even below the detection limits. Due to the diversity of galaxy properties (i.e., PAH fraction or FIR-radio correlation scatter), some of these sources will be missed in our 24+radio prior list, but might have detectable SPIRE or (sub)mm fluxes.

We carry out a “blind” search (i.e., unguided by priors) for additional sources by running SExtractor (Bertin & Arnouts 1996) on the residual images. We combine any new sources detected this way with the previous list of priors, and then re-run *galfit* to see if the new, blindly selected sources can be detected.

We limit our blind-extraction region to fall within the 24+radio catalog coverage. This is done by measuring the perimeter of the 24+radio sources and setting pixels outside of the perimeter to not-a-number (NaN) values. Then we run SExtractor on the masked residual image with a detection threshold of about 2.5–3.0 (depending on the band, with a GLOBAL background) to detect all possible residual sources. We subject the resulting SExtractor catalogs to careful visual inspection. We run *galfit* again, fitting over all previously selected prior sources in that band plus the new sources from blind detection that pass the visual inspection. The new residual source candidates are finally kept and included in our catalog if their final S/N is above 3, typically only 1/4 of the initial SExtractor candidates.

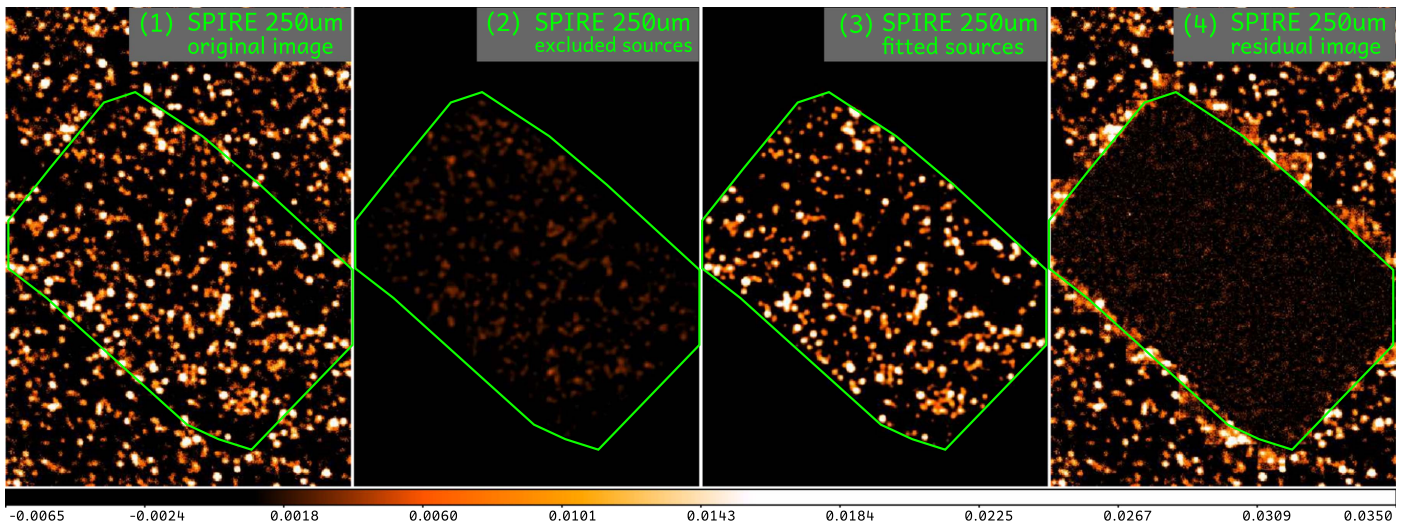
We illustrate this blind detection in Figure 7, using SPIRE 350  $\mu\text{m}$  data. Panel (1) and panel (2) show the same residual image before extracting additional sources, and the green circles in panel (2) highlight the additional sources which are extracted by SExtractor and then measured with *galfit* to have  $S/N > 3$ . Panel (3) shows the new residual image after we run the final *galfit* fitting, including prior sources and the additional sources that were retained. We list the number of final detected additional sources at each band in Table 1.

We have decided not to include these additional sources in the scientific analysis presented in the final sections of this paper for a number of reasons. Their positions are uncertain, and in most cases it is not trivial to associate them with optical/near-IR counterparts. Also, a certain fraction of these blindly detected sources are probably spurious, often a superposition of a number of fainter sources simulating a brighter object. It is unlikely that these detections are residuals from poor fits to brighter sources, as this is rarely seen in the FIR/mm residual images. Nevertheless, there are also likely genuine  $z > 3$  sources among these objects—for example, the  $z = 5.3$  submillimeter galaxy HDF850.1 (Walter et al. 2012) is the additional source ID 850011 in our additional source catalog.

## 4. An Example: Optimizing the Prior Source List around GN20

Figure 8 illustrates the procedure of choosing prior sources for fitting based on their SED-predicted fluxes for the fairly crowded field around the luminous  $z = 4.055$  starburst GN20 (Pope et al. 2006; Daddi et al. 2009b), or ID 564 in our catalog. The circles in the figure represent all 24+radio prior sources





**Figure 6.** SPIRE 250  $\mu\text{m}$  images, shown as an example of our photometry method. Panel (1) is the original SPIRE 250  $\mu\text{m}$  image. The green boundary indicates the sky coverage of our 24+radio catalog (Section 2.4). Panel (2) is the modeled image of the sources that were excluded from the super-deblended prior list (Section 3.2), based on their fluxes predicted by SED fitting. These faint sources were not used for PSF fitting, while all other sources are fitted as priors because of their higher likelihood of being detected at 250  $\mu\text{m}$ . Panel (3) shows the best fitting model image of these prior sources after running *galfit* photometry (e.g., Section 2.2), and panel (4) is the resulting residual image. The image in panel (1) is the sum of panels (2)–(4). The intensity scales are the same for all panels, as indicated by the bottom color bar. The flux density unit is Jy/beam.

(Section 2), including those *selected* (green) and *excluded* (orange) from fitting.

As described in the previous sections, for each band we first construct the SED for each 24+radio prior source using the photometry available so far, and then choose which objects to fit in the current image and which to exclude. Here we consider four marked sources in Figure 8 as examples: ID 532, 564, 658, and 659. ID 532 is one of the brightest sources at 16  $\mu\text{m}$  and 24  $\mu\text{m}$  bands, but becomes relatively fainter in the SPIRE bands. It has a spectroscopic redshift of 0.9746, and thus the redder SPIRE wavelengths fall longward of the FIR SED peak. The extracted SED is shown in the last panel of Figure 9. Based on SED fitting at wavelengths shorter than 500  $\mu\text{m}$ , the SED-predicted 500  $\mu\text{m}$  flux density of ID 532, augmented by twice the uncertainty, falls below the cutoff threshold:  $S_{\text{SED}} + 2\sigma_{\text{SED}} < S_{\text{cut},500\mu\text{m}}$ , where the  $S_{\text{cut},500\mu\text{m}}$  is the critical value we chose, as shown in the fifth panel of Figure 5. Thus we do not fit ID 532 at 500  $\mu\text{m}$ . Meanwhile, its predicted flux contribution to the observed 500  $\mu\text{m}$  image is subtracted as described in Section 3.3, so that we can measure the real, deboosted flux for the remaining prior sources.

ID 658 and ID 659 are two fainter sources near GN20 that have the same spectroscopic redshift  $z = 4.055$  (GN20.2a, GN20.2b respectively; they are part of a proto-cluster at that redshift). Tan et al. (2014) presented photometry and SEDs for these objects, as well as mm continuum and spectral line observations. The mm photometry provides tight constraints on the Rayleigh–Jeans tail of the dust SED. In Figure 9, we compare our deblended photometry and SEDs with the photometry of Tan et al. (2014). They did not measure the SPIRE fluxes for GN20, and their flux uncertainties at PACS 100 and 160  $\mu\text{m}$  are larger than ours. Their mm photometry agrees very well with our SEDs, noting that the SEDs are fitted only with the black data points in this work. For the fainter GN20 proto-cluster members, ID 658 and ID 659, Tan et al. (2014) provided only PACS fluxes, which all have  $S/N < 3$ . Our photometry provides more data points, and our SEDs are

in very good consistency with the previous mm continuum observations.

## 5. Simulation-based Flux and Uncertainty Corrections

We use extensive Monte Carlo simulations in our super-deblended photometry for two main purposes. First, simulations can verify potential flux biases, so that we can correct for any imperfect sky background or other subtle systematic effects. Second, they help us understand and calibrate the uncertainties of photometric measurements.

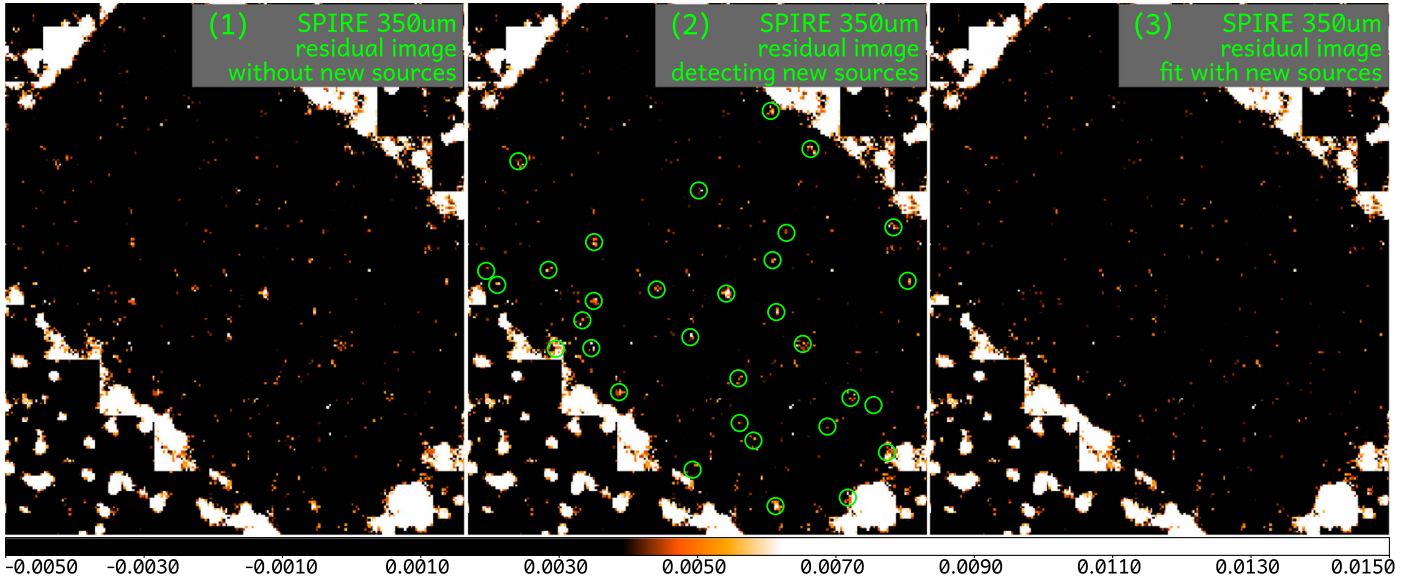
In fact, the flux uncertainties estimated by *galfit* are based on diagonalizing and projecting the covariance matrix, and hence may not properly reflect real errors (Peng et al. 2002). In simulations, we have not only the *galfit* output but also the simulated input information. Therefore we can link the real uncertainty to the *galfit* output with measurable source properties. We will present the Monte Carlo simulation method in Section 5.1, the flux bias analysis in Section 5.2, and the flux uncertainty analysis in Section 5.3.

### 5.1. Monte Carlo Simulation

We simulate one source at a time in the faint-source-subtracted image.<sup>31</sup> The simulated source positions are generated randomly within the survey area, and do not avoid other real, bright sources. Their input flux densities,  $S_{\text{in}}$ , are uniformly distributed in log space within the range from  $\sim 1\sigma$  to  $\sim 30\sigma$ , where  $\sigma$  is the mean flux density uncertainty at each band (see Table 1).

We add each extra simulated source to the 24+radio prior list, which has already been filtered for that band (i.e., with all selected sources, see Section 3.2, as well as additional sources detected in the residual images, see Section 3.3), and then perform the PSF fitting photometry including all active priors and an extra prior at the position of the simulated source. We

<sup>31</sup> The faint-source-subtracted image is the image with excluded source models subtracted from the original image; see Section 3.3.



**Figure 7.** Here we show the extraction of additional sources from the residual image. Panels (1) and (2) show the residual image after fitting and subtracting sources from the catalog of priors. The two images are identical, but panel (2) shows the new sources detected by SExtractor. Panel (3) is the new residual image after re-running the photometric fitting including these additional sources. All panels have the same intensity scale, and the flux density unit is Jy/beam.

also perform two-pass *galfit* fitting as described in Section 2.2 to allow variation in the positions of priors in the second pass if they have high S/N in the results of the first pass. At the end of each pass, we run the corrections described in the following Sections 5.2 and 5.3. The *galfit* output flux density of the simulated source is  $S_{\text{out}}$ , and the *galfit* flux density uncertainty is  $\sigma_{\text{galfit}}$ .

Finally, this process is repeated  $\sim 3000$  times. Therefore, for each band, we have  $\sim 3000$  values for  $S_{\text{in}}$ ,  $S_{\text{out}}$ , and  $\sigma_{\text{galfit}}$ . In addition, we measured several additional properties for each simulated source: for example, the local rms noise value  $\sigma_{\text{rms noise}}$  for each source, measured in the instrument noise image data; the local absolute flux density in the residual image (hereafter  $S_{\text{residual}}$ ), measured by computing the total absolute pixel values in the PSF aperture for each fitted prior source; the local scatter in the residual image, measured also in the PSF circle; and the *crowdedness* parameter, computed by summing up the Gaussian weighting of all sources at current source  $i$  position,

$$\sum_{j=1}^N e^{(-d_{j,i}^2/d_{\text{PSF}}^2)}, \quad (1)$$

where  $d_{j,i}$  is the angular distance in arcsec from source  $j$  to current source  $i$  and  $d_{\text{PSF}}$  is the FWHM in arcsec of the PSF for the band being analyzed. Defined in this way, the *crowdedness* is a weighted measure of the number of sources present within the beam, including the specific source under consideration.

These measurable parameters are designed to provide key information about the quality of fitting and the local crowding (hence blending) of prior sources. We will make use of them with the simulation results in order to calibrate the best possible flux bias and flux uncertainty corrections.

In addition to the uncertainties estimated in this way, we add in quadrature to each object's flux uncertainty the appropriate contribution (when relevant) to account for astrometric dispersion when fitting at fixed spatial positions. This term is estimated by computing the dispersion in measured fluxes for

bright sources at fitted position versus those at the positions optimized in the fit.

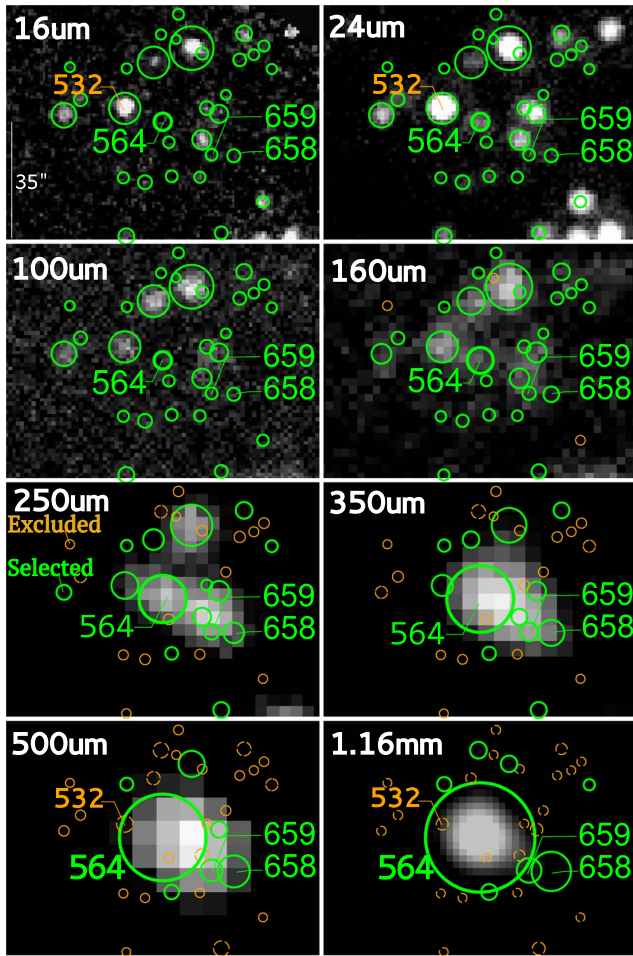
## 5.2. Flux Bias Correction

The statistics of the differences between input and output fluxes,  $S_{\text{in}} - S_{\text{out}}$ , can be used to verify and correct the bias of the *galfit* flux measurements. In Figure 10 we present the analysis of  $S_{\text{in}} - S_{\text{out}}$  based on simulation data for SPIRE 350  $\mu\text{m}$  as an example. In panel (1),  $S_{\text{in}} - S_{\text{out}}$  is plotted against the simulation input flux density  $S_{\text{in}}$ , which covers the range  $3\text{--}12\times$  in terms of the detection limit  $\sigma_{350\mu\text{m}}$ . The overall median of  $S_{\text{in}} - S_{\text{out}}$  is generally small, indicating that the (constant) background levels used in the *galfit* work are quite accurate.

For all bands we explore the dependence of  $S_{\text{in}} - S_{\text{out}}$  on the following three key properties: (1) the flux uncertainty  $\sigma_{\text{galfit}}$ , which is a direct output of *galfit* fitting; (2) the residual flux  $S_{\text{residual}}$ , measured in the output residual image; and (3) the *crowdedness*, which indicates prior source blending as described in the previous section. When testing parameters (1) and (2), we further use the rms noise  $\sigma_{\text{rms noise}}$  to normalize these two parameter values in order to account for the local rms variation in case of non-uniform depth of the data over the whole GOODS-N region. For example, if sources at the map center have smaller rms noise than sources near the edges, the outer sources will likely have higher  $\sigma_{\text{galfit}}$  and  $S_{\text{residual}}$  simply because of their locations.

As shown in panels (2)–(4) of Figure 10, we bin simulated sources by each of the key properties. In each bin we compute the median of  $S_{\text{in}} - S_{\text{out}}$ , which is just the flux bias  $S_{\text{bias}}$ . Then we fit  $S_{\text{bias}}$  in each bin by a three-order polynomial function of the  $X$  axis parameter. The fitted polynomials are shown as the red curves in panels (2)–(4). Note that the quantities along the  $X$  axes are measurable for each real fitted source; thus these functions are also applicable to each real source.





**Figure 8.** Multi-wavelength cutouts around the bright submillimeter galaxy GN20 (Pope et al. 2006; Daddi et al. 2009b), which is ID 564 in this work. Each circle represents a 24+radio prior source (Section 2). Green circles are selected for fitting (marked “Selected” in the 250  $\mu\text{m}$  panel), while orange circles are excluded (marked “Excluded”). The circles have sizes proportional to the source flux densities for each band (i.e., the measured fluxes for fitted sources, or the SED-predicted fluxes for excluded sources, but with a minimum size of  $0''.5$  and a maximum size equal to the PSF FWHM). The SEDs of the four marked sources, ID 564 (GN20), 658 (GN20.2a), 659 (GN20.2b), and 532 are shown in Figure 9. The SED-predicted flux density for ID 532 (plus twice the uncertainty) is too faint at 500  $\mu\text{m}$  (as indicated by the vertical dashed line in the SED figure) and hence is excluded from the fitting list in this band.

### 5.3. Flux Uncertainty Correction

The Monte Carlo simulations also provide a path toward obtaining reliable estimates of flux uncertainties. The flux error returned by *galfit* ( $\sigma_{\text{galfit}}$ ) only reflects the fitting uncertainty in a formal way (e.g., Peng et al. 2002). In fact, the *galfit* formalism is optimized for optical/near-IR images and assumes uncorrelated noise (all pixel are independent) in the data. Thus it cannot fully account for real uncertainties, especially in the presence of correlated noise and confusion noise. Imperfect flux uncertainties will strongly affect the validity of SED fitting for individual sources and the assessment of detections or lack thereof.

We analyze the statistics of  $(S_{\text{in}} - S_{\text{out}})/\sigma_{\text{galfit}}$  with simulation data, which should have a rms dispersion of 1.0 by definition if uncertainties are well defined and Gaussian-like. Taking the 350  $\mu\text{m}$  simulation as an example, in the left upper panel of Figure 11 we bin simulated objects by the measurable parameter ( $\sigma_{\text{galfit}}/\sigma_{\text{rms noise}}$ ) as in previous section, and in the middle and right upper panel we show the histograms of

$(S_{\text{in}} - S_{\text{out}})/\sigma_{\text{galfit}}$  (in blue) and  $(S_{\text{in}} - S_{\text{out}})/\sigma_1$  (in red), where  $\sigma_1$  is the first-step corrected flux uncertainty. The width of the blue histogram is clearly much broader than 1.0, indicating that the values  $\sigma_{\text{galfit}}$  are substantially underestimated.

We determine correction factors as a function of the  $X$  axis parameter, requiring that, in bins defined for quantities in the  $X$  axis, the rms dispersion of  $(S_{\text{in}} - S_{\text{out}})/\sigma$  is equal to 1.0. In each bin, we compute the rms of  $(S_{\text{in}} - S_{\text{out}})$  and the median of  $\sigma_{\text{galfit}}$ , and then calculate their ratio, which is exactly the factor that we need for correcting  $\sigma_{\text{galfit}}$  to the real uncertainty of  $(S_{\text{in}} - S_{\text{out}})$ . As in the previous section, we fit the correction factor as a three-order polynomial function against the  $X$  axis parameter. The fitted polynomial functions are shown as the red curves in the figure. In this way, we obtain the first-step corrected  $\sigma_1$ , whose histogram is shown in red, and is much closer to a width of 1.0.

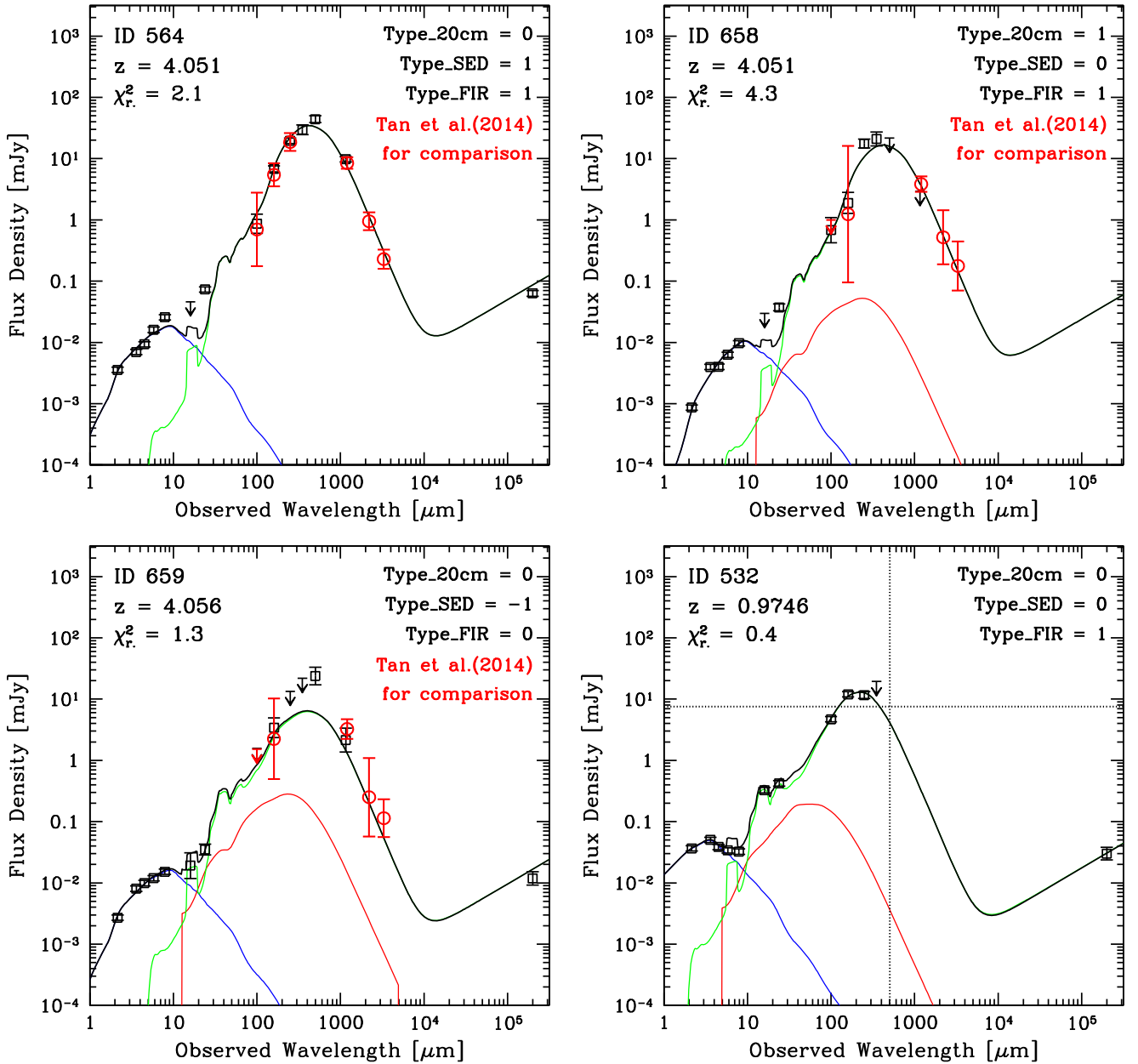
The first-step corrected uncertainty histograms can sometimes have asymmetric shapes, which indicate that there are still some systematic biases. Thus we take the first-step corrected uncertainty  $\sigma_1$  as input and perform the second and the third steps. In the second step, we first perform the flux bias correction with parameter  $S_{\text{residual}}/\sigma_{\text{rms noise}}$ , and then perform the uncertainty correction with the same parameter as illustrated by the middle row panels of Figure 11. Then, similarly, in the third-step we apply the flux bias correction, followed by the flux uncertainty correction with the parameter *crowdedness*. Therefore, with these correction recipes, we are able to apply the correction to real fitted sources.

We have also tried multiple iterations or different combinations of the three steps. For example after the three-step processing, we use the output corrected flux and uncertainty as the input for several more iterations of the three-step processing. However, we find that the corrected values and histograms are stable and cannot be further improved in this way.

Similar figures for all examined bands are reported in the Appendix B (Figures 46–52). Generally, for objects whose *galfit* uncertainties are around or below the median for that band, we find that the flux uncertainties are often quite underestimated, while for objects with the largest uncertainties (typically those affected by substantial deblending uncertainties), the values are often overestimated (i.e., the real errors are smaller as judged from our simulations). An analogous effect is seen for the *crowdedness* parameter. We have analyzed a number of other parameters (including flux itself, spatial positions, redshift, different kind of *crowdedness* and residual definitions, etc.), and we found no other significant dependency to further correct flux uncertainties.

## 6. Super-deblended Photometry—Overall Work Flow

Based on the simulation analyses, we apply the correction recipe to real fitted sources. For each fitted source, we begin with the flux uncertainty given by *galfit* fitting ( $\sigma_{\text{galfit}}$ ), then measure the local noise on the rms image ( $\sigma_{\text{rms noise}}$ ), the local residual flux on the residual image ( $S_{\text{residual}}$ ), and the *crowdedness* parameter using all fitted prior source coordinates at each band. Therefore the aforementioned three measurable parameters can be obtained for each fitted source. We compute the correction factor from each parameter and correct flux bias and uncertainty sequentially in the steps described in Sections 5.2 and 5.3. Finally, we obtain the super-deblended GOODS-N FIR+mm photometry results.



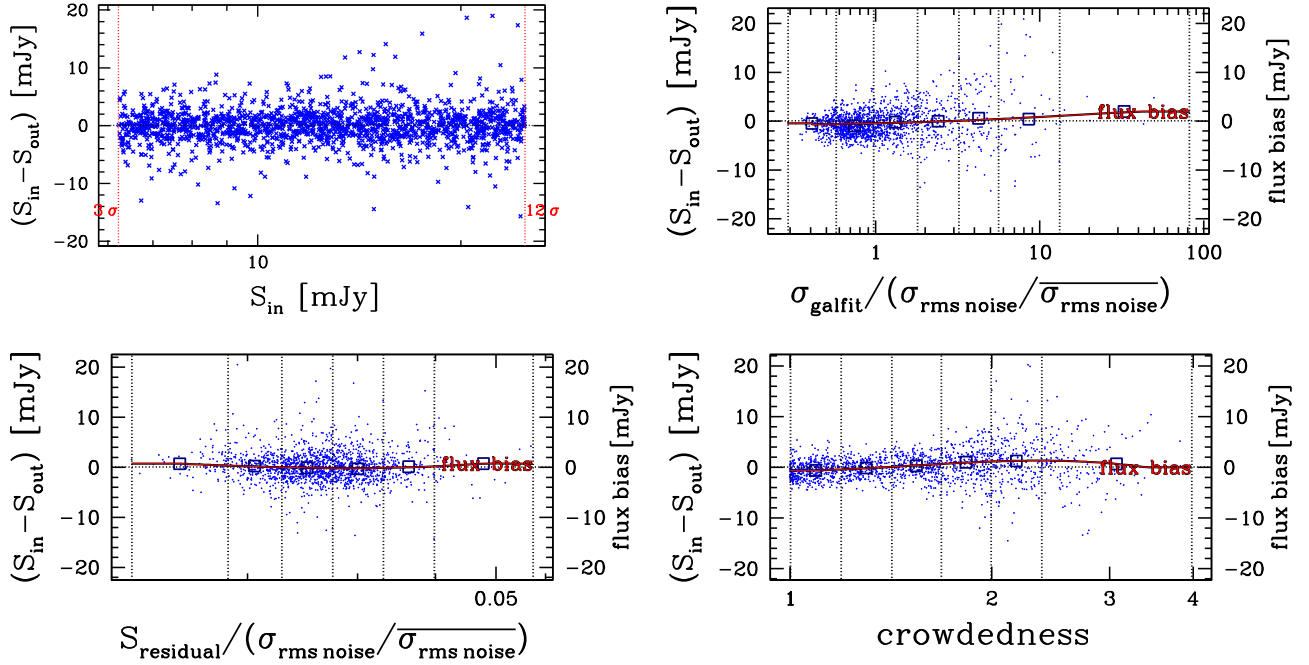
**Figure 9.** SEDs of the four sources marked in Figure 8. In the panels for ID 564 (GN20), 658 (GN20.2a), and 659 (GN20.2b), we show the red data points from Tan et al. (2014) for comparison. These are not used in the SED fitting, but show very good consistency with our measurements. For ID 532, the  $500 \mu\text{m}$  SED-predicted flux density (indicated by the vertical dashed line) becomes a fainter cutoff threshold (the horizontal dashed line), even when increased by twice the uncertainty, and hence this source is excluded from fitting at this band. In the upper-left corners we show the redshift and best fitting SED reduced  $\chi^2$  ( $\chi^2_r$ ), which is computed over the wavelength range 100–1100  $\mu\text{m}$ . The three SED tuning parameters are shown in the upper right corners. See Section 3.1 for details.

Our super-deblended method of optimizing prior source lists with the help of the optimized SED fitting ensures that we are fitting all the most likely detectable prior sources for each band. The flux and uncertainty corrections we derive on a source-by-source basis make our detections and SED fitting more reliable. We list the final mean flux density uncertainty ( $1\sigma$ ) and the number of  $S/N \geq 3$  sources ( $N_{S/N \geq 3}$ ; i.e., single-band detected catalog sources) at each band in Table 1.

When all photometric measurements have been derived for all bands, we run a final pass of global SED fitting from which we derive various galaxy properties. While a detailed discussion of extraction of physical parameters and their

uncertainties from FIR galaxy SEDs (which would include AGN components, dust temperature, IR8, radio excess, etc.) is beyond the scope of this work, for the sake of the scientific analysis presented in this paper (see Section 8), it is worth mentioning the derivation of the best fitting  $L_{\text{IR}}$ , and also SED-driven photometric redshifts. The latter are in most cases a simple refinements of optical/near-IR photometric redshifts, corresponding to the redshift within the allowed range (close to the previous optical/near-IR photometric redshifts) where the best fit to the FIR+mm SED is obtained. However, for objects with no previous optical/near-IR photometric redshift, we fit over the whole range  $0 < z < 8$ , and the SED fitting provides





**Figure 10.** Panel (1) shows the difference between input and output flux densities of simulated sources ( $S_{\text{in}} - S_{\text{out}}$ ) vs. the input (i.e., true) flux density for SPIRE 350  $\mu\text{m}$  Monte Carlo simulation. In panels (2)–(4), we bin simulated sources by three measurable parameters: the *galfit* flux density uncertainty normalized by the local rms noise, the flux density on residual image also normalized by the local rms noise, and the *crowdedness* parameter which measures the blending situation at each source position (see text in Section 5.2). Dashed vertical lines indicate the bins. We measure the mean of  $S_{\text{in}} - S_{\text{out}}$  (blue square data points), which is just the flux bias, and then fit a polynomial function to link the flux bias to each parameter (red curves). In this way, bias can be better corrected for different sources. Similar figures for all other bands are shown in [Appendix](#).

an estimate of their photometric redshift and uncertainty. In a future publication we will investigate the possible derivation of FIR/mm/radio photometric redshifts and their performance, improving over the work of Daddi et al. (2009b) and several others.

The overall flow chart of our super-deblended method is presented in Figure 12.

## 7. Quality Checks and Known Limitations

### 7.1. Comparing SED Predictions with Measurements

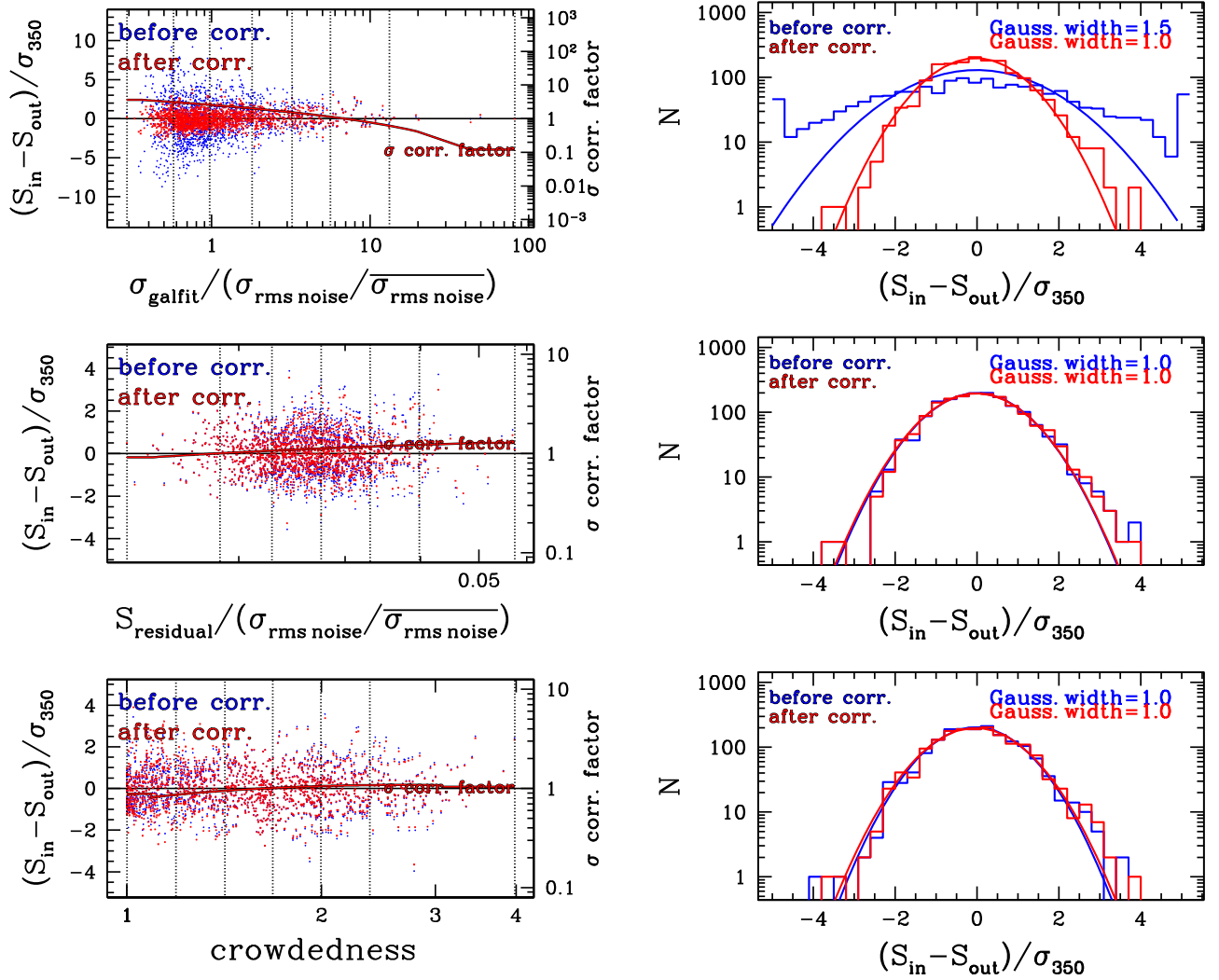
An important step in our procedure is to predict fluxes for sources in a given band before actually performing measurements in that data set. Sources with predicted fluxes (plus twice flux prediction uncertainty) brighter than  $S_{\text{cut}}$  are retained to be fitted with *galfit*. For those fitted sources, we can now compare their measured fluxes to the SED predictions in order to investigate the presence of any potential bias in those predictions (and, by extension, also the predictions for the faint/excluded/subtracted sources).

These comparisons are shown in Figure 13. In the left panels we compare  $S_{\text{predict}}$  and  $S_{\text{galfit}}$ , and in the right panels we show the difference ( $S_{\text{galfit}} - S_{\text{predict}}$ ) divided by the total uncertainty, plotted versus  $S_{\text{predict}}$ . The error bars represent the combined uncertainties of  $\sigma_{\text{galfit}}$  and  $\sigma_{\text{predict}}$ .

$S_{\text{predict}}$  and  $S_{\text{galfit}}$  generally agree well with each other. There is a small fraction of sources with  $S_{\text{predict}}$  lower than  $S_{\text{galfit}}$  at the faint end, as seen in the left panels (but less obvious in the right panels when actual flux uncertainties are taken into account). This small bias is driven by the requirement to have a significant *galfit* detection in order to be shown in the plot, and vanishes when this selection bias is taken into account.

### 7.2. Goodness of SED Fitting and Measurement Errors

Another important overall check of our procedure is evaluating the goodness of fit of the global SEDs that we derived with our adopted template libraries, to verify at the same time that our templates are appropriate and that the error bars are sensible. We concentrate here on the FIR+mm range between 100 and 1100  $\mu\text{m}$ , which is the range most affected by blending that we are interested to test, and also because mid-IR and radio deviations from our SEDs might be expected because galaxies can be more complicated than the models (AGN components with different shapes/slopes, strength of PAHs, etc.). As we have already shown with examples of SEDs (see, e.g., Figure 9), the reduced  $\chi^2$  values are generally reasonable and close to unity. We verified this aspect more systematically for the whole sample, considering the overall IR detection S/N ratio (see Equation (2) in Section 8). For the 518 galaxies with  $S/N_{\text{FIR+mm}} > 10$ , we find a median  $\chi_r^2 = 1.0$ , while a median  $\chi_r^2 = 0.8$  is found for the  $5 < S/N_{\text{FIR+mm}} < 10$  sources. About 2% of all galaxies in either sample cannot be fit by our models, with  $\chi_r^2 > 4$ . In most cases this appears to be due to an erroneous or problematic redshift (photometric or spectroscopic), either due to an error or because of blending/superposition between foreground and background objects, because of gravitational lensing or simply line-of-sight alignment. In some cases it is quite clear that only part of the FIR/mm emission is due to the galaxy under consideration as the IRAC prior, and that an additional component comes from a higher redshift object that contributes additional flux at longer wavelengths. Many of these could be scientifically interesting cases (e.g., lenses or the high redshift background galaxies).



**Figure 11.** Left panels: for simulated objects (at  $350\ \mu\text{m}$  as an example), we show the difference between input (real) and output (measured) flux divided by flux uncertainty in  $Y$  axis (i.e.,  $(S_{\text{in}} - S_{\text{out}})/\sigma$ ). From top to bottom, we analyze this quantity against three parameters as indicated by the  $X$  axis label. We bin objects as indicated by the vertical dashed lines and compute the rms in each bin for deriving the correction factors as described in Section 5.3 (i.e., imposing that, in bins defined for quantities in the  $X$  axes, the rms dispersion of  $(S_{\text{in}} - S_{\text{out}})/\sigma$  is equal to 1.0). We show the data points before and after correction (i.e.,  $(S_{\text{in}} - S_{\text{out}})/(\sigma, \text{uncorrected})$  and  $(S_{\text{in}} - S_{\text{out}})/(\sigma, \text{corrected})$ ) in blue and red, respectively, and with a solid line representing a best fitting Gaussian profile to the inner part of each histogram. Right panels: the histograms of the  $(S_{\text{in}} - S_{\text{out}})/\sigma$  in logarithm space. After the three-step corrections, the histograms are well-behaved and generally well consistent with a Gaussian distribution. Similar figures for all other bands are shown in Appendix B.

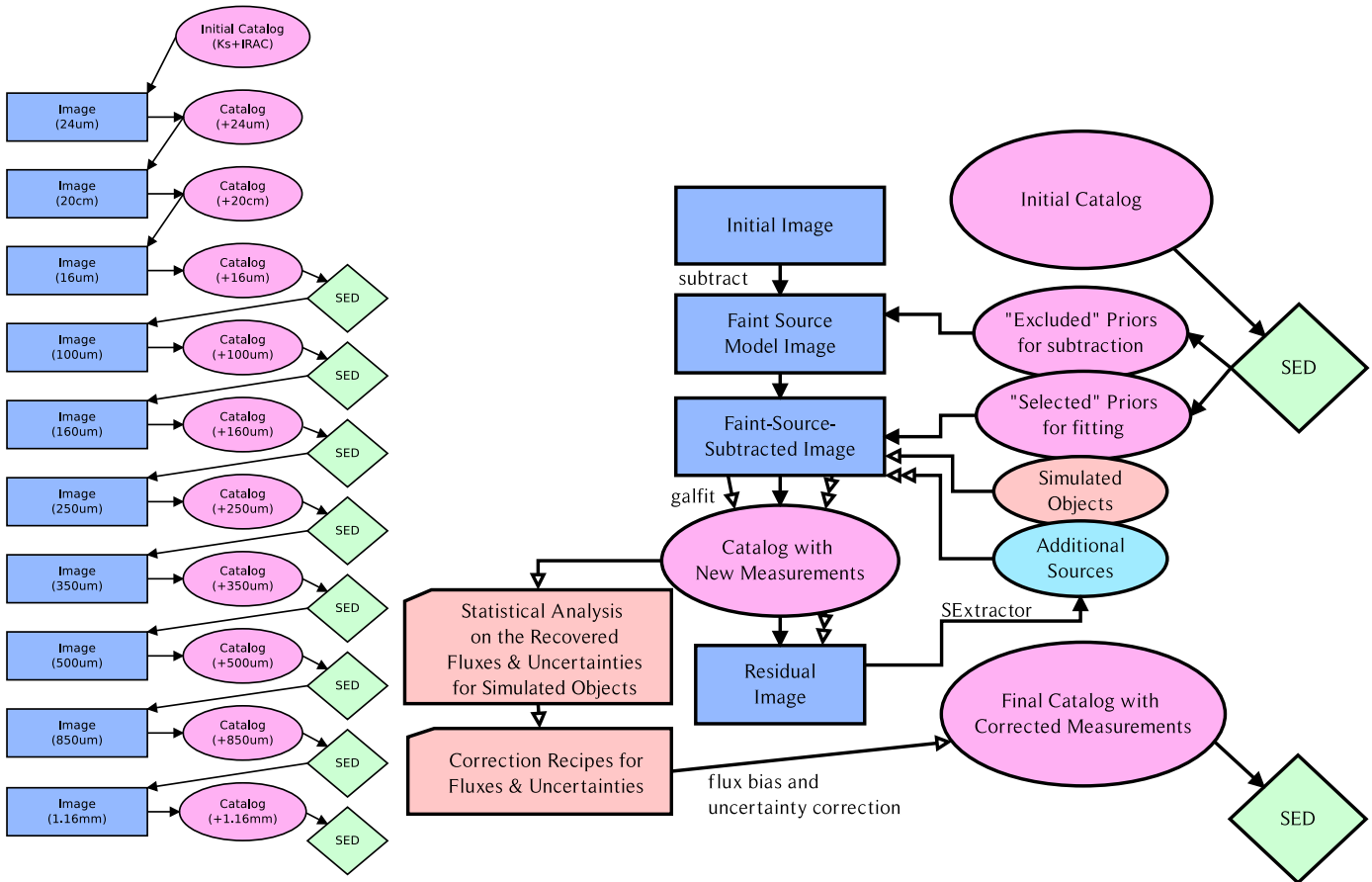
### 7.3. Comparison to Literature Results

Here we compare our  $24\ \mu\text{m}$ , PACS, and SPIRE flux measurements and uncertainties with those of Elbaz et al. (2011) and HerMES (DR3; Roseboom et al. 2010),<sup>32</sup> and our  $1.16\ \text{mm}$  catalog with that of Penner et al. (2011).

In Figure 14, we compare our  $24\ \mu\text{m}$  flux and uncertainty histograms with the Elbaz et al. (2011) catalog. In the left panel, the flux histogram of  $S/N_{24\ \mu\text{m}} \geq 3$  sources in this work is shown in blue, and that of Elbaz et al. (2011) is shown in red. The shaded area under each histogram indicates the sources with FIR+mm multiband combined  $S/N_{\text{FIR+mm}} \geq 5$  and  $S/N_{24\ \mu\text{m}} \geq 3$ , where  $S/N_{\text{FIR+mm}}$  is defined as per Equation (2) in Section 8.1. Elbaz et al. (2011) do not include (sub-)mm bands (i.e., for their sources the combined  $S/N$  only involves  $100\text{--}500\ \mu\text{m}$  photometry). For the bright sources, the flux histograms from both works are consistent, while for

fainter sources, the flux histogram from this work shows more prior sources. We have about 450 more  $24\ \mu\text{m}$  detected candidate prior sources for the longer-wavelength photometry than those used in Elbaz et al. (2011). The shaded area, or the FIR to mm and  $24\ \mu\text{m}$  detected sources, also has a larger number of detections in our catalog. The newly detected sources are fainter. Notice how IR-detected sources fall off very rapidly below  $100\ \mu\text{Jy}$ , suggesting that our prior sample is quite complete and possibly redundant for *Herschel* sources at typical redshifts (although we might still be somewhat incomplete for the highest redshifts, as previously noted in Section 2.4). The right panel shows flux uncertainties, corrected based on simulations (see Section 5.3). They are generally higher than those in Elbaz et al. (2011), but should have less bias and be more representative of the true uncertainties in the measurements (see Sections 5.3 and B). For the FIR and  $24\ \mu\text{m}$  detected sources in the shaded area, our flux uncertainties are also generally higher.

<sup>32</sup> <http://hedam.lam.fr/HerMES/>



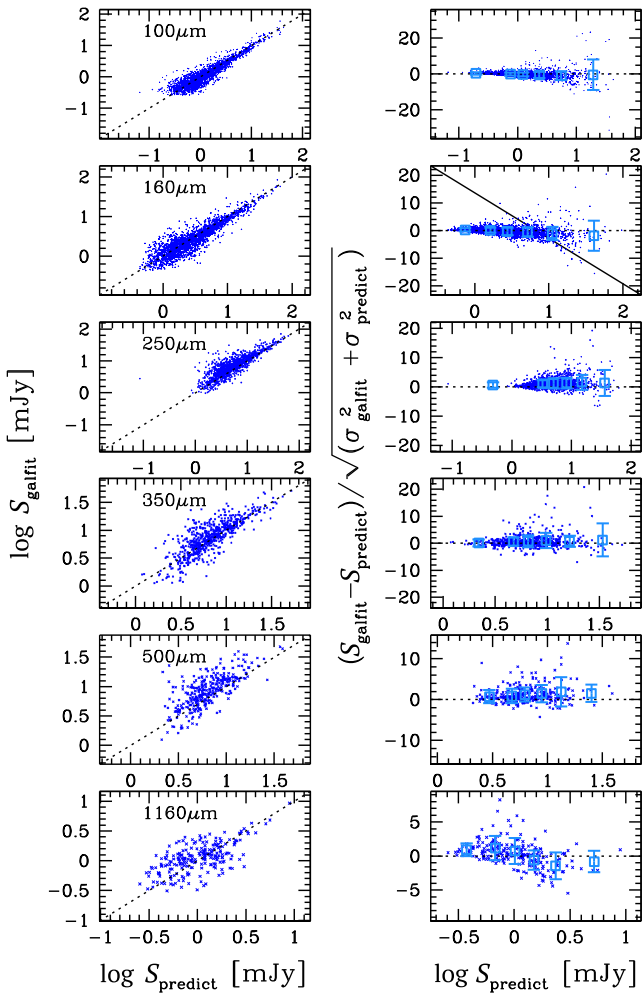
**Figure 12.** Left: flow chart of our photometry procedures as applied to all bands. Catalogs are indicated by purple ellipses, SED fitting processes by green diamonds, and images by blue rectangles. From top to bottom, following the arrows, we start with the IRAC catalog and perform photometry at  $24\ \mu\text{m}$ ,  $20\ \text{cm}$ , and  $16\ \mu\text{m}$ , as described in Section 2. Then SED fitting is used to help determine an updated source prior list for each FIR/mm band, as discussed in Section 3. Right: super-deblended procedures at each FIR/mm band. For each source in the initial catalog, we run SED fitting and predict the flux and uncertainty in the current band (Section 3.1). Then we split prior sources into two lists: “priors to subtract” and “priors to fit” (Section 3.2). We make a source-model-image for sources in the former list and subtract it from the initial observed image, resulting in a faint-source-subtracted image on which the *galfit* photometry is then performed for sources in the latter list (Section 3.3). We further blindly extract additional sources in the residual image and re-run the *galfit* photometry with “priors in residual” (Section 3.4), which results in the “catalog with new measurements.” Meanwhile, we generate simulated objects in the faint-source-subtracted image, and then recover fluxes and analyze and calibrate uncertainties as described in Section 5. We derive “correction recipes” that make use of measurable parameters; thus for each fitted source, we can measure these parameters and apply the corrections to real measurements. Finally, we obtain the “catalog with corrected measurements” for each band and fit the SED again, which will be used in the next band or as the final output after the last FIR/mm band has been analyzed.

In Figure 15, we compare the  $100\ \mu\text{m}$  flux and uncertainty histograms from this work to those from the catalog of Elbaz et al. (2011). In the left panel, our work has about 200 more sources and extends to a slightly fainter flux range. In the right panel, after applying the simulation-based flux uncertainty correction (see Section 5.3), our flux uncertainties are lower than those in Elbaz et al. (2011). This is due both to the greater number of fainter sources in the left panel, and to our simulation-based corrections to the flux uncertainties.

Other PACS and SPIRE bands have similar histograms to that at  $100\ \mu\text{m}$  and hence are not shown here. The comparisons of our  $1.16\ \text{mm}$  flux measurements and uncertainties with those of Penner et al. (2011) are shown in Figure 16. We use their deboosted fluxes when possible. The number of detected sources is small at this wavelength, but our catalog has 13 more detections. The sources span a similar range of flux in both works. After applying the simulation-based flux uncertainty corrections (see Section 5.3), the flux uncertainties in this work are slightly higher than those in Penner et al. (2011). Similar to the case at  $24\ \mu\text{m}$ , we believe that the difference is mainly due

to the fact that our corrections yield more reasonable flux uncertainties (based on the simulation results).

The one-to-one flux comparisons for the three SPIRE bands and at  $1.16\ \text{mm}$  are presented in Figures 17 and 18. We cross-match common sources (within a limiting separation of  $1''$ ) and compare the fluxes from this work to those from Elbaz et al. (2011) or the HerMES catalog. The majority of the sources show consistent fluxes, but some sources in the Elbaz et al. (2011) catalog have higher SPIRE fluxes, which is due to source blending. The two most extreme outliers are marked with their IRAC IDs in the first panel. They are ID 14896 (ID 2132 in Elbaz et al. 2011) and ID 10611 (ID 1512 in Elbaz et al. 2011). ID 14896 is one of very few sources that have higher fluxes in our catalog than in that of Elbaz et al. (2011). This is due to the complex blending situation; Elbaz et al. (2011) assign most of the flux to a nearby source but little to this object. Figure 19 shows the multiband image cutouts around this source. The three sources, ID 14896, ID 14867, and ID 15027, become blended at  $160\ \mu\text{m}$  and longer wavelengths. Their SEDs are shown in Figure 20. For comparison, the red data points report values from Elbaz



**Figure 13.** Comparison between SED-predicted flux densities ( $S_{\text{predict}}$ ) before the *galfit* photometry step, and the newly measured flux densities ( $S_{\text{galfit}}$ ) after the *galfit* photometry step for each band, as a demonstration of the performance of our SED prediction. The left panels show the flux-flux comparison in log space, while the right panels show flux vs. flux difference weighted by the combined error in linear space, for the same bands as the left panels. Because the SED flux and uncertainty ( $S_{\text{predict}}$  and  $\sigma_{\text{predict}}$ ) values are used (a) to determine which sources are too faint to be detectable, and (b) to subtract their flux contributions from the observed data (see Section 3.3), it is essential to ascertain that SED predictions are not subject to strong biases. In this figure,  $S_{\text{predict}}$  and  $S_{\text{galfit}}$  are generally consistent in all bands, although with larger dispersion (indicated by the error bars in the right panels) at longer wavelengths. The small skewness in the results can be explained by the fact that only significant ( $>3\sigma$ ) detections are shown, which produces a bias toward sources with higher  $S_{\text{galfit}}/S_{\text{predict}}$  ratios.

et al. (2011) for ID 14896 and ID 15027. ID 14867 has no counterpart in Elbaz et al. (2011). In the third panel, ID 15027, with a relatively low spectroscopic redshift  $z = 1.371$ , is not likely to have SPIRE flux as bright as the red data points. In the first panel, the 20 cm and 500  $\mu\text{m}$  fluxes of ID 14896 suggest that its 250  $\mu\text{m}$  flux is not likely to be as low as that reported by Elbaz et al. (2011). Thus, in our opinion, for this case our deblending leads to more reasonable results. For the other extreme case mentioned above (ID 10611 and its surrounding sources), we have also checked their image cutouts and SEDs. The blending situation is very similar to that described above, and in the catalog of Elbaz et al. (2011) most of the SPIRE flux is attributed to ID 10611 (e.g., like ID 15027). Thus the detailed figures and SEDs are not shown here.

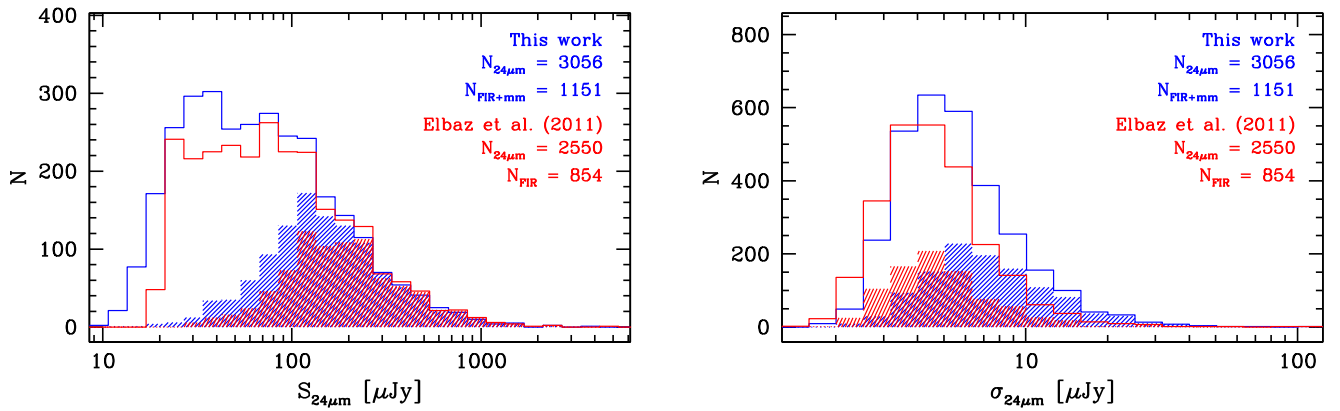
Similar considerations apply to the HerMES catalogs (see Figure 18). In this case, we note that there are large differences with our measurements at 350 and 500  $\mu\text{m}$ , substantially larger than the uncertainties from both catalogs. Generally, we find that many sources in the HerMES catalog have their flux densities overestimated, likely because of unresolved blending. The number of single-band detected sources is almost doubled in our catalog compared to the HerMES catalog, for all SPIRE bands. The flux density uncertainties in HerMES catalog (using confusion noise computed from the inverted covariance matrix and the PSF matrix; see Roseboom et al. 2010, 2012) tend to be larger than ours.

#### 7.4. Effects of Varying the Prior Density

Our method attempts to use all available information on galaxies in the region of sky under consideration in order to define an ideal set of priors to fit FIR/(sub)mm data, using the smallest possible number of useful sources. Increasing the number of priors will increase photometric errors for all galaxies on average because of increased blending. This effect is exemplified in Figure 21, where we show the normalized flux errors for the 100 and 250  $\mu\text{m}$  bands as a function of the *crowdedness* parameter. Despite these two bands having quite different numbers of priors per beam (0.3 and 1, respectively; see also Figure 1), the normalized behavior of the flux error versus *crowdedness* is similar for the two bands. To first order, the relative flux error scales proportionally to *crowdedness*, although a second order polynomial fit is a better description for values of *crowdedness*  $> 2$ . At 250  $\mu\text{m}$ , as a result of the much larger beam, we are fitting a smaller absolute number of priors (see Figure 21 and Table 1), which on average have higher values of *crowdedness*. The median value for the *crowdedness* is 1.64 (1.05) at 250  $\mu\text{m}$  (100  $\mu\text{m}$ ), and the median relative flux error is similarly larger 1.63 (versus 1.02). The increase of the typical “noise” is the price to pay for fitting all reasonable priors. We note that this effect is not so dramatic, and errors are still quite acceptable, even if we are dealing with a density of priors approaching 1 per beam (in possible contradiction with results and discussions in Karim et al. 2013; Scudder et al. 2016).

Figure 5 shows that, given the information in hand, it is not possible to reduce much further the number density of priors without failing to fit objects that could be potentially fairly bright in the FIR/mm imaging data. For example, raising  $S_{\text{cut}}$  from 3 to  $6\sigma$  would reduce  $\rho_{\text{beam}}$  only from 1 to  $\sim 0.8$ . On the other hand, if we were to reduce  $S_{\text{cut}}$  to  $1.5\sigma$ , we would reach  $\rho_{\text{beam}}$  of about 2. It is interesting at this point to evaluate what the impact of these choices would be. Figure 22 shows the distribution of *crowdedness* values for the different cases. Reducing the source density from 1.0 to 0.8 per beam has a fairly minor effect on the implied *crowdedness*, and hence on the noise. Increasing the source density much beyond 1 has a much stronger effect: the fraction of sources in reasonable isolation, with *crowdedness*  $< 1.5$ , drops from 42% ( $\rho_{\text{beam}} = 1.0$ ) to 15% ( $\rho_{\text{beam}} = 2.0$ ), while the sources with poor measurements due to blending with *crowdedness*  $> 2.5$  rise from 12% ( $\rho_{\text{beam}} = 1.0$ ) to 41% ( $\rho_{\text{beam}} = 2.0$ ). Therefore, if we were to accept more sources for fitting, including some of those below  $S_{\text{cut}} = 3\sigma$ , this would result in a significantly increase in the noise and poorer performance for all objects. Staying within the limit  $\rho_{\text{beam}} \lesssim 1$ , as we have done for SPIRE in this paper, appears to be a reasonable choice.





**Figure 14.** Comparison of histograms of the 24  $\mu\text{m}$  flux densities (left) and flux density uncertainties (right) between this work and that of Elbaz et al. (2011). The open areas of the histogram show all prior sources, while the shaded areas represent the sources that are significantly detected (with  $S/N_{\text{FIR+mm}} \geq 5$  for our sources, or  $S/N_{\text{FIR}} (100 \text{ to } 500 \mu\text{m}) \geq 5$  for Elbaz et al. 2011 sources; see the definition of  $S/N_{\text{FIR+mm}}$  in Section 8.1 and Equation (2)). Results from this work are shown in blue, and for Elbaz et al. (2011) in red. In the left panel, the histograms of bright sources are consistent, but there are more sources from this work at fainter flux densities. This means that we have about 450 more candidate prior sources for the longer-wavelength photometry than in Elbaz et al. (2011). The FIR to mm detected sources, which are the red and blue shaded areas, also show consistency for brighter fluxes but are more numerous in this work. In the right panel, due to the simulation-based flux uncertainty correction (see Section 5.3), the uncertainties in this work are generally higher, even for the FIR and 24  $\mu\text{m}$  detected sources in the shaded area. The  $N_{24\mu\text{m}}$  is the number of  $S/N_{24\mu\text{m}} \geq 3$  sources in each catalog, and  $N_{\text{FIR+mm}}$  is the number of FIR-to-mm (or FIR 100–500  $\mu\text{m}$  for Elbaz et al. 2011 catalog) sources.

### 7.5. Robustness to Source Subtraction Errors

We are subtracting faint-source models from the images before running the photometric analysis. This procedure is prone to errors, as we might be over- (or under-) estimating the flux values to subtract. We expect that any such error would be automatically taken into account by our simulation-based approach. For example, if we were over-subtracting the fluxes of many faint sources, this would create, to first order, a negative (false) background, which we would detect and correct with our bias measurements, which in practice define the effective background in the data. To second order, the over-subtracted sources in this example, after zero-level determination, would result in increased background fluctuations and noise, because they are not likely to be spatially homogeneous, and they will have a variety of flux values. This extra noise would be reflected in our simulation-based calibration of uncertainties and thus taken into account in our error bars.

### 7.6. Known Limitations

We are aware of a number of possible remaining problems that might still affect our flux measurements and their uncertainties.

One is the presence of additional sources (most often at high redshifts), which could significantly contribute flux in the data but which are not in our prior list. Such objects could spuriously boost the inferred flux (and hence lead to underestimated errors) of other priors. Judging from our simulations, this effect does not seem to be significant, since the distribution of  $(S_{\text{in}} - S_{\text{out}})/\sigma$  closely resembles a Gaussian with width equal to 1 (see Figure 11 and Appendix B).

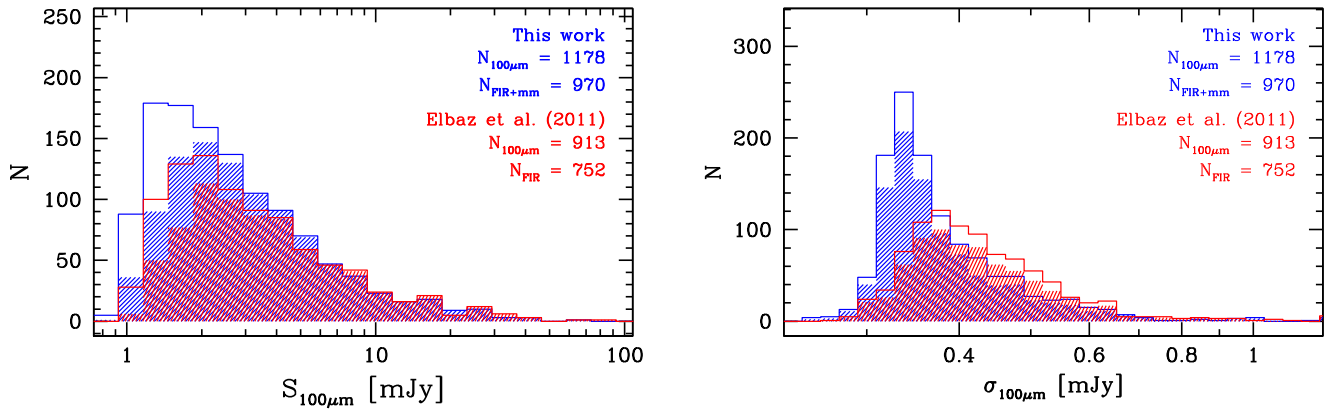
Also, a fraction of photometric redshifts or even spectroscopic redshifts could be incorrect. Experience from various surveys have shown that a few percent of galaxies may have substantially incorrect redshift measurements or estimates, even when using values classified as reliable, as we do here. Note that in order to accept a spectroscopic redshift from the literature, we require that it is in agreement within 10% in  $(1+z)$  with the photometric redshift, to avoid possible wrong redshifts. Using a wrong redshift can affect SED predictions. To some extent, allowing flexibility in dust temperatures

(or, more specifically, in  $\langle U \rangle$  values) moderates the impact because of the redshift-temperature degeneracy, especially if the redshifts are not catastrophically wrong. In this case, due to SED flux prediction uncertainties, there could be a few objects that are retained for *galfit* fitting instead of being dropped. This would modestly increase *crowdedness*, but this is probably a negligible effect. Some objects might be erroneously dropped from fitting while still being bright enough for detection in the data. The latter cases could give rise to sources that would be detectable in the residual images if they are isolated. If they fall atop or near other priors, they may corrupt their photometry, biasing recovered fluxes high. Of course, in all cases, the impact of wrong redshifts would extend beyond simple photometric accuracy to the derivation of wrong bolometric  $L_{\text{IR}}$  and SFR, thus affecting the science.

Finally, another limitation is that we are using *galfit* with a constant background. Angular variations in the background would act as increased noise. Carefully accounting for background variations could improve photometric uncertainties, but is limited by our knowledge of PSF wings, and of course by the very large beams in the SPIRE and (sub)mm imaging. We attempted a correction of background variations at 24  $\mu\text{m}$  and 16  $\mu\text{m}$ , where the beam is still fairly small and manageable. Dealing with this for PACS and SPIRE imaging is beyond the scope of the present work.

#### 7.6.1. Some Considerations about Simulation Methodology

We carry out our simulations (Section 5) independently for each band, adding objects randomly into the real image data with many Monte Carlo repetitions. An alternative approach, which we have not yet employed, might simultaneously and consistently simulate source fluxes over their full SEDs, then create fully simulated deep survey images in all of the bands, then carry out a full photometric analysis on these simulated data. We recall, however, that we are also performing flux measurements one band at a time, and assigning uncertainties to them in a reliable way at each band, as verified by our simulations. We use band-to-band information only to select which galaxies to fit and which to exclude, and we do this in a very conservative way because, in any case, we fit many more



**Figure 15.** Comparison of histograms of the 100  $\mu\text{m}$  flux densities (left) and flux density uncertainties (right) between this work and that of Elbaz et al. (2011). Colors and histograms are defined in Figure 14. In the left panel, this work has about 200 more sources and spans a slightly deeper range of flux densities. In the right panel, after applied the simulation-based flux uncertainty correction (see Section 5.3), the uncertainties in this work are lower than those in Elbaz et al. (2011) catalog, both because of the larger number of fainter sources in the left panel and the simulation-based corrections to the flux density uncertainties.

galaxies down to fainter fluxes than we can actually measure. Errors or limitations of this procedure would eventually factor into the uncertainty budget of the flux measurements, which is inferred from our simulations. Therefore it appears that simultaneous modeling of galaxies in all bands, while perhaps useful, might not be a critical limitation of this work in its current version.

However, we recognize that there is interesting additional information that could be obtained by a full multiband image simulation. In particular, such a simulation could permit an estimate for the amount of correlations among flux uncertainties between adjacent wavelengths, induced by the flux correlation of the galaxy SEDs themselves and their slowly rolling shapes (regardless of whether the signal is from primary galaxies or from a superposition of those in the background; e.g., from confusion). Neglecting this might actually spuriously enhance the significance of source detection, as performed in the next sections by coadding the S/N over several bands. We use a fairly conservative threshold to limit the impact of this problem. We will return to this issue with an appropriate statistical approach in future work.

Also, a full multiband image simulation, including clustering (e.g., Béthermin et al. 2017) and extended to very faint galaxies, could allow us to address the issue of the nature of “additional sources” (Section 3.4; i.e., to have information on which fraction of them might be spurious superpositions of several faint galaxies). This is also left to future work.

## 8. Results and Discussion

In this final section we use the FIR+mm catalog developed in this work to study the IR emission of high redshift galaxies and to provide an estimate of the SFR density in the universe, up to high redshift.

In order to evaluate whether a source is detected in the FIR+mm catalog, we adopt a simple approach considering the combined S/N over PACS, SPIRE, 850  $\mu\text{m}$ , and 1.16 mm bands:

$$\begin{aligned} (S/N)_{\text{FIR+mm}}^2 &= (S/N)_{100\mu\text{m}}^2 + (S/N)_{160\mu\text{m}}^2 \\ &+ (S/N)_{250\mu\text{m}}^2 + (S/N)_{350\mu\text{m}}^2 + (S/N)_{500\mu\text{m}}^2 \\ &+ (S/N)_{850\mu\text{m}}^2 + (S/N)_{1.16\text{mm}}^2. \end{aligned} \quad (2)$$

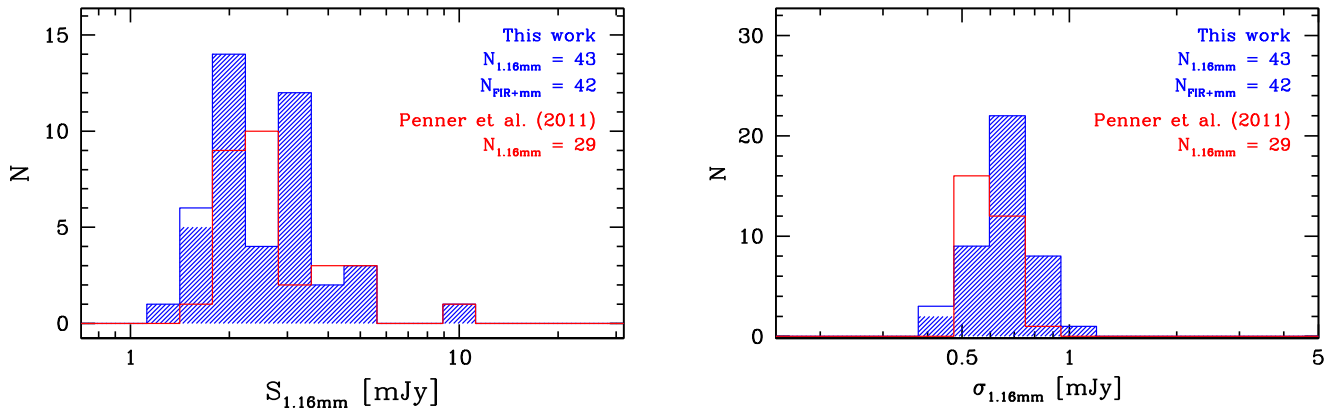
Our  $(S/N)_{\text{FIR+mm}}^2$  definition is equivalent to a Chi-squared field whose expectation value (in the presence of pure noise) for 7 bands being added is equal to 7, and the probability to have  $S/N_{\text{FIR+mm}} \geq 5$  is  $7 \times 10^{-4}$  for pure Gaussian random fields (Bloomfield et al. 2016). In such an ideal case (but see Section 7), this would produce at most a couple of spurious IR detection from our starting pool of prior positions. Among the 3306 24+radio sources, 1109 are detected with  $S/N_{\text{FIR+mm}} \geq 5$ , and 518 are very significantly detected with  $S/N_{\text{FIR+mm}} \geq 10$ .

While we have applied our photometric technique to the whole GOODS-N area, in some of the remainder sections (e.g., for the derivation of the SFR density, Sections 8.2–8.9), we focus our analysis on objects that lie in the central 134 arcmin<sup>2</sup> region (hereafter *goodArea*), avoiding the outer perimeter where the instrumental noise at 24  $\mu\text{m}$  starts to rise, and hence the 24  $\mu\text{m}$  prior sources will be less complete. This *goodArea* contains 862  $S/N_{\text{FIR+mm}} \geq 5$  sources and 427  $S/N_{\text{FIR+mm}} \geq 10$  sources.

### 8.1. Redshift, SFR, and Stellar Mass Distributions

In Figure 23, we show the SFR versus redshift for all 3306 24+radio sources in the full GOODS-N field. The SFR are derived from the pure dust component IR luminosity over 8–1000  $\mu\text{m}$ , excluding any AGN torus components as derived from our SED fitting (see Section 3.1). Among all the detected sources (henceforth, the FIR+mm sources), 106 sources have  $z \geq 2.5$  (FIR+mm SED based photometric redshift or spectroscopic redshift when available), 70 have  $z \geq 3$ , and 20 have  $z \geq 4$ . Within the *goodArea*, we find 71 sources at  $z \geq 2.5$ , 49 at  $z \geq 3$ , and 14 at  $z \geq 4$ .

In Figure 23 we indicate the SFRs of typical MS galaxies at three fiducial stellar masses ( $M_* = 10^{12} M_\odot$ ,  $10^{11} M_\odot$ , and  $10^{10} M_\odot$  respectively), computed with the Equation (A1) of Sargent et al. (2014). This FIR+mm sample has a non-uniform distribution in both SFR and stellar mass. At lower redshift (e.g.,  $z < 1$ ), the limits of the SFR and stellar mass both rise quickly. At  $z < 1$ , the GOODS-N FIR+mm data can detect emission from MS galaxies with  $M_* < 10^{10} M_\odot$ , but at  $z > 1$  all detected galaxies are more massive. At higher redshift (e.g.,  $z > 3$ ), the sample is clearly biased toward high SFRs and stellar masses ( $M_* > 10^{11} M_\odot$ ). Both the effective limits of SFR and stellar mass have risen by one to two orders of magnitude compared to those at  $z \sim 1$ .



**Figure 16.** Comparison of histograms of the 1.16 mm flux densities (left) and flux density uncertainties (right) between this work and that of Penner et al. (2011). Colors and histograms are defined in Figure 14. The flux densities from Penner et al. (2011) are deboosted values when available. Most 1.16 mm-detected sources are also FIR+mm-detected sources. In the left panel, the present work reports significant detections for about 10 more sources over a similar range in flux density. In the right panel, after applying the simulation-based correction (see Section 5.3), the flux density uncertainties in this work are slightly larger than those in the Penner et al. (2011) catalog.  $N_{1.16\text{ mm}}$  is the number of  $S/N_{1.16\text{ mm}} \geq 3$  sources in each catalog, and  $N_{\text{FIR+mm}}$  is the number of sources with PACS 100  $\mu\text{m}$  to 1.16 mm  $S/N \geq 5$  and  $S/N_{1.16\text{ mm}} \geq 3$ .

We present our FIR+mm catalog in Table 2. The full machine-readable version can be accessed via the online journal; here we only show a few important columns whose descriptions are in the caption.

### 8.2. UV Unattenuated SFRs

In order to estimate the total SFRs in our IR-detected galaxies, we need to add the unattenuated SFR that can be directly recovered in the rest-frame UV ( $\text{SFR}_{\text{UV}}$ ). We thus cross-matched the 3D-*HST* catalog for the rest-frame UV flux densities of the FIR+mm galaxies ( $S/N_{\text{FIR+mm}} \geq 5$ ). A total of 94% of them have rest-frame 1400 Å flux densities from the 3D-*HST* catalog. The  $\text{SFR}_{\text{UV}}$  is in general a small contribution (10% typically) compared to  $\text{SFR}_{\text{IR}}$  for our IR selected sample. Figure 24 shows the ratio of  $\text{SFR}_{\text{UV}}$  and  $\text{SFR}_{\text{IR}}$  versus stellar masses and redshifts (and in *goodArea*). The left panel shows sources with rest-frame 1400 Å flux densities and stellar masses, colored by their redshifts. The right panel shows all FIR+mm sources (also in *goodArea*), with different colors indicating the method adopted to estimate  $\text{SFR}_{\text{UV}}$ , as discussed below. Method (1): if a galaxy is in the 3D-*HST* catalog (matched within 1''0 radius) and has rest-frame 1400 Å flux densities, then we use the Kennicutt (1998) calibration:  $\text{SFR}_{\text{UV}} = 1.4 \times 10^{-28} \times 4\pi \left( \frac{d_L}{\text{cm}} \right)^2 \left( \frac{S_{1400\text{Å}}}{\text{erg s}^{-1} \text{cm}^{-2} \text{Hz}^{-1}} \right)$ . We also check that using instead the rest-frame 1700 Å flux densities (also provided by 3D-*HST*) leads to no more than 0.1 dex difference in average. Method (2): if a galaxy has no rest-frame 1400 Å but has a stellar mass, then we use a simple correlation between  $\text{SFR}_{\text{UV}}/\text{SFR}_{\text{IR}}$  and stellar mass (as shown in the left panel of Figure 24, analogous to the stellar mass-attenuation relation that is fairly constant with redshift; Pannella et al. 2015) that we obtained with the galaxies in method (1):  $\log_{10}(\text{SFR}_{\text{UV}}/\text{SFR}_{\text{IR}}) = 13.3 - 1.4 \log_{10}(M_*)$ . This relation appears only to be valid for  $\log_{10}(M_*) \geq 9.6$ , and does not hold at very low stellar masses, where IR-detected objects are mainly SB sources (e.g., see the right panel of Figure 25). For the only two sources with  $\log_{10}(M_*) < 9.6$  and no UV continuum, we just adopted the average  $\text{SFR}_{\text{UV}}/\text{SFR}_{\text{IR}}$  value of the  $\log_{10}(M_*) < 9.6$ , UV-continuum-available sample ( $\text{SFR}_{\text{UV}}/\text{SFR}_{\text{IR}} = 0.22$ ). Finally, for eight galaxies that do

not have UV continuum nor stellar mass, as well as for a small sample of  $\sim 28$  objects where UV continuum is most likely dominated by the emission of an AGN (e.g., indicated by the mid-IR AGN SED fits and optical/near-IR images), we adopted the median  $\text{SFR}_{\text{UV}}/\text{SFR}_{\text{IR}}$  of the whole sample ( $=0.11$ ). In the following sections, we add  $\text{SFR}_{\text{UV}}$  and  $\text{SFR}_{\text{IR}}$  to obtain the total SFR ( $\text{SFR}_{\text{UV+IR}}$ ) for each galaxy, and use the total SFR for all further analysis.

### 8.3. Stellar Mass and SFR Properties

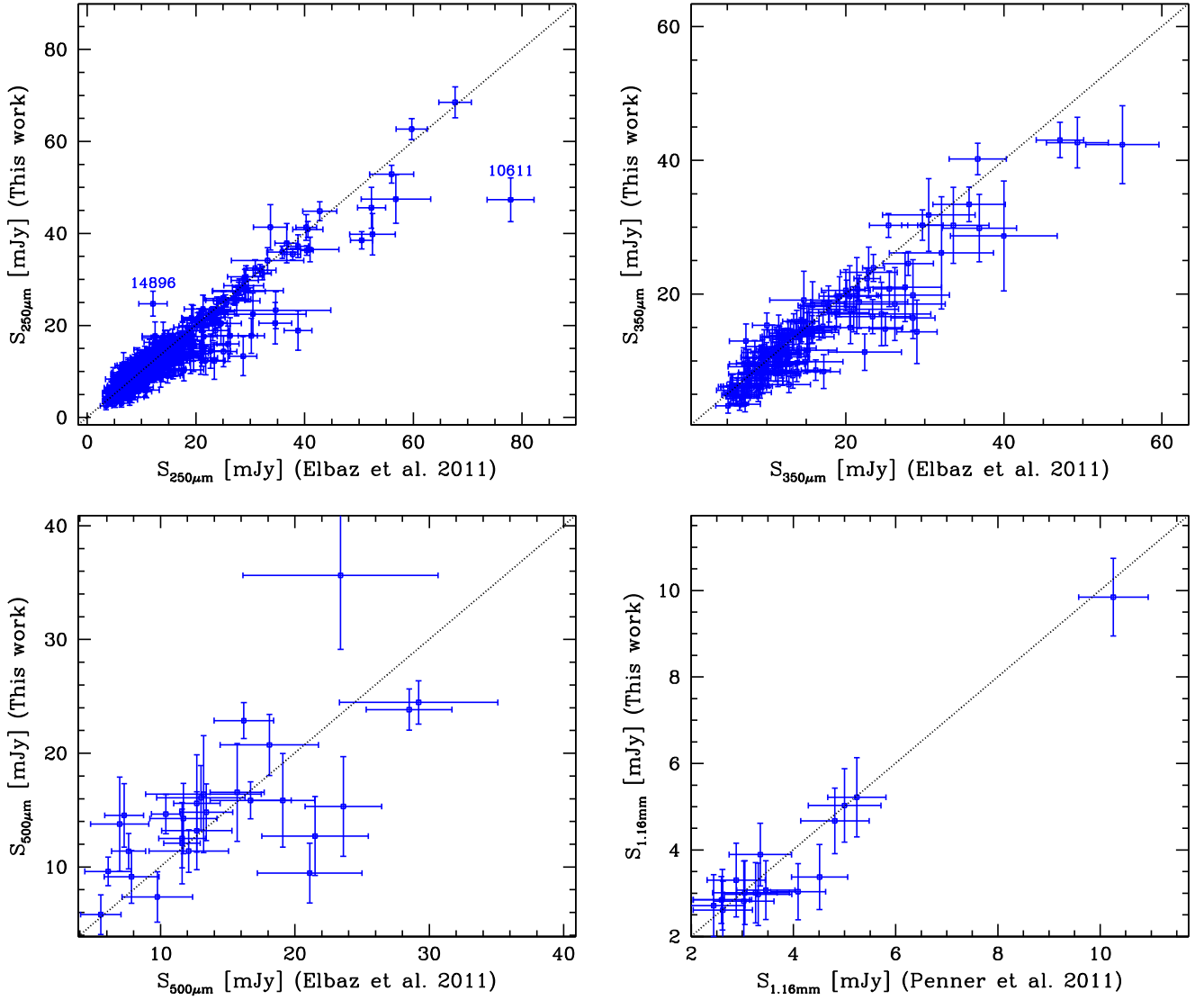
In Figure 25 we compare our sample with empirical MS correlations in redshift bins up to  $z = 6$ . Only *goodArea* FIR+mm sources with stellar masses from 3D-*HST* or Pannella et al. (2015) are shown here; seven FIR+mm sources lack stellar mass estimates and are not shown. The differences between the three MSs considered here (from Sargent et al. 2014; Béthermin et al. 2015; Schreiber et al. 2015) are relatively large at both  $z < 1.5$  and  $z > 4$ , but the different estimates are more consistent at  $1.5 < z < 4$ . In the highest redshift bins, the Sargent et al. (2014) MS has the lowest SFR normalizations. The difference between these three MS relations (in some cases their extrapolations) can reach a factor of three at  $M_* \sim 10^{11.5} M_\odot$  and  $z \sim 4-6$ , where global uncertainties and lack of direct IR data in the previous derivations still prevent a solid assessment of the MS level.

The red symbols in each panel indicate sources classified as strong starbursts during the SED fitting (Section 3.1), relative to the MS of Sargent et al. (2014). They are relatively rare, and mostly have masses  $M_* \lesssim 10^{10.5} M_\odot$ . The fraction of SBs in this sample is 9.8% at  $1 < z < 4$ , intermediate between the fraction of 30%–50% for submillimeter galaxy-like selection at the highest luminosities and the fraction of 2%–3% derived for a sample at  $1.5 < z < 2.5$  selected by stellar mass (complete to  $M_* \geq 10^{10} M_\odot$ ; Rodighiero et al. 2011<sup>33</sup>).

The shaded area in each panel indicates the stellar mass range over which the sample is incomplete, evaluated by analyzing SMFs for star-forming galaxies as described in the next section. Our FIR+mm sample should have a high degree of completeness at the characteristic stellar mass  $M_{\text{star}}^*$  in the

<sup>33</sup> Their definition of SB is the same as that which we have adopted in this work (i.e.,  $\log \text{SFR} - \log \text{SFR}_{\text{MS}} > 0.6$  dex).





**Figure 17.** Comparisons between flux density measurements in this work and literature catalogs at 250, 350, 500  $\mu\text{m}$ , and 1.16 mm. Literature fluxes for the three SPIRE bands are taken from Elbaz et al. (2011), and the literature 1.16 mm flux densities are from Penner et al. (2011). The figures show flux comparisons for sources detected in both our catalog and in previous work and cross-matched based on position. Elbaz et al. (2011) used PACS-detected sources as priors for SPIRE photometry, while in this work we use SED fitting to the near-infrared to PACS, SPIRE, and radio photometry to determine the prior sources for SPIRE flux extraction, and we subtract the faint-source flux contribution from the observed data. Therefore our flux density measurement tend to be smaller than the literature values, especially when a source is blended with neighbors. Two most extreme outliers are marked with their IRAC ID in the first panel. These are ID 14896 (Elbaz et al. 2011, ID 2132) and ID 10611 (Elbaz et al. 2011, ID 1512). We discuss ID 14896 in detail in Section 7.3 and the next figures. Although the offset for ID 10611 has the opposite sense, we have found this source has a very similar blending issue.

four lowest redshift bins, where  $M_{\text{star}}^*$  for star-forming galaxies is estimated to be in the range of  $10^{10.3}-10^{10.9} M_{\odot}$  at  $0.2 < z < 2.5$  according to Davidzon et al. (2017), Ilbert et al. (2013), and Muzzin et al. (2013).

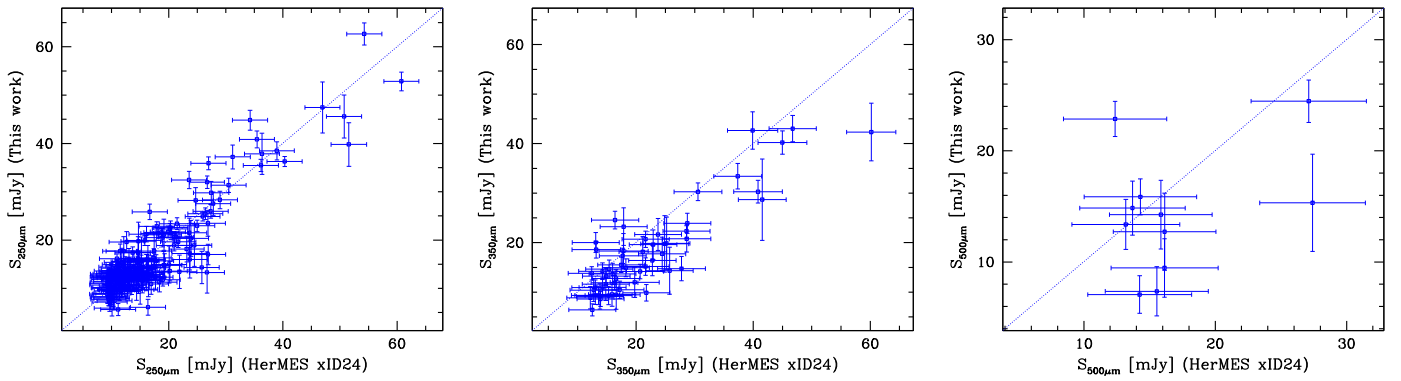
At  $z > 3$ , estimates of  $M_{\text{star}}^*$  fall in the range  $10^{10.74-11.56} M_{\odot}$  (Ilbert et al. 2013; Muzzin et al. 2013; Grazian et al. 2015; Davidzon et al. 2017). At these redshifts, the FIR+mm sample can only probe a small number of galaxies at the high-mass end of the SMF.

#### 8.4. Comparing with Literature SMFs

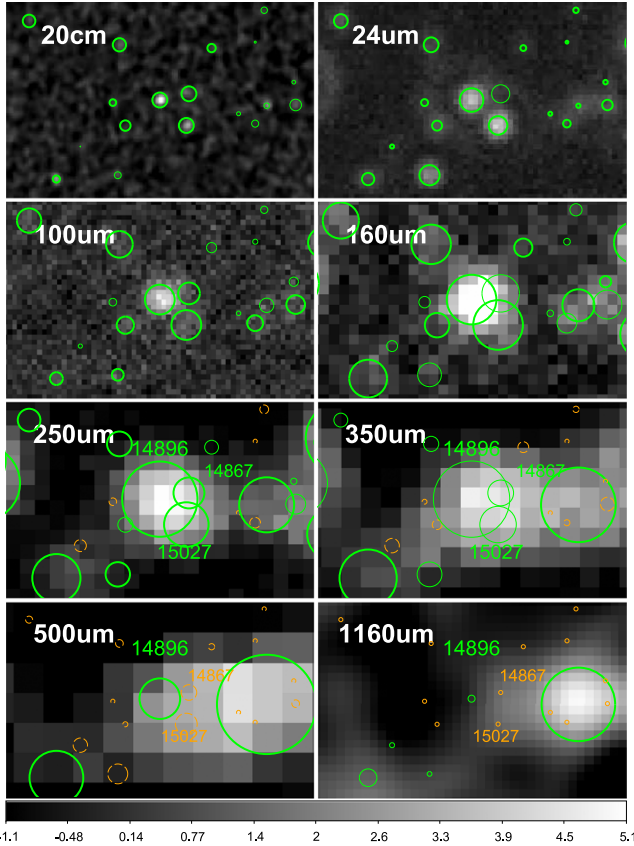
To evaluate the incompleteness of our FIR+mm sample we consider sources in each redshift bin, further dividing them into bins of stellar masses in Figure 26, and compare their

distributions with various star-forming galaxy SMFs from the literature.

The literature SMFs are convolved to a common stellar mass uncertainty ( $\sigma_{M_*}$ , see labels in each panel) in order to account for measurement errors when comparing with our observed stellar mass distributions (blue and red histograms in the figure, representing the full FIR+mm sample and the SB sub-sample respectively). This is done by first using the Eddington bias-corrected (i.e., intrinsic) SMFs in these literature, then convolving them with a common  $\sigma_{M_*}$  (which is redshift-dependent), following the approach described in Appendix A of Ilbert et al. (2013; except for Song et al. 2016 where the authors have not made that correction). We multiply by the *goodArea* co-moving volume to convert SMF (in units of



**Figure 18.** Flux density measurement comparisons for sources with  $S/N \geq 3$  in both this work and the HerMES catalog (DR3; Roseboom et al. 2010). The three panels show SPIRE 250, 350, and 500  $\mu\text{m}$  measurements, respectively.



**Figure 19.** Multi-wavelength cutouts around the outlier ID 14896, which has a significantly lower flux in our catalog than in that of Elbaz et al. (2011). The discrepancy arises because Elbaz et al. (2011) assign more flux to ID 15027 than in our catalog. Each circle represents a 24+radio prior source (Section 2). As in Figure 8, green circles are the fitted prior sources, while orange circles are the excluded ones. The circle sizes scale with flux (see caption of Figure 8). The SEDs of the three heavily blended sources, ID 14896, ID 14867, and ID 15027, are shown in Figure 20.

number per volume) to the absolute number of galaxies within each stellar mass bin and each redshift bin.

The literature SMFs do not always agree with each other, especially at the low-mass end and at high redshifts, but in general they agree with our observed histograms up to the  $3 < z < 4$  bin without doing any fitting. Muzzin et al. (2013) presented two best-fits to the SMF at each redshift: one with

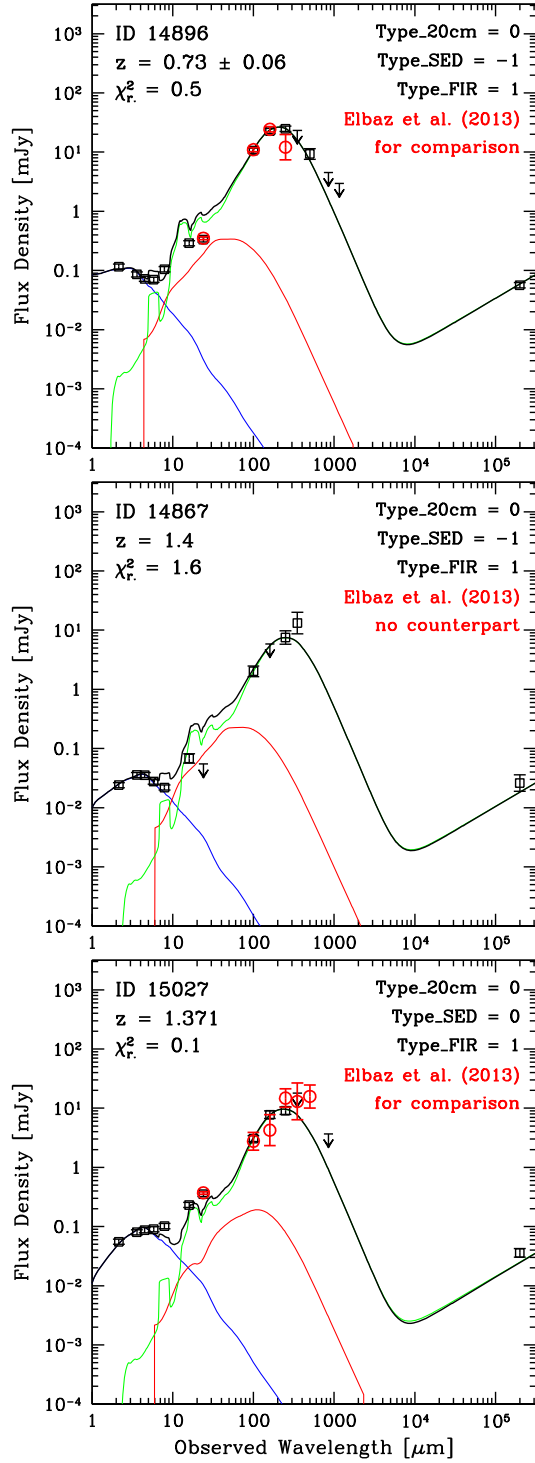
fixed slope at the low-mass end, the other with non-fixed slope. Their fixed-slope SMFs agree better with the other two  $z < 4$  SMFs in the literature, while the non-fixed-slope SMFs show large discrepancies at the low-mass end compared to other SMFs at  $z > 2$ . In order to probe the variety of SMFs, we show their non-fixed-slope SMFs in the figures. Note that the SMF of Davidzon et al. (2017) does not fully cover our highest redshift bin. We have linearly extrapolated their SMFs at  $3.5 < z < 4.5$  and  $4.5 < z < 5.5$  to our bin at  $5 < z < 6$  for purpose of comparison. Because their SMF is even higher at  $4.5 < z < 5.5$  than at  $3.5 < z < 4.5$ , the extrapolated SMF is also higher.

To evaluate the stellar mass range (shown as the shaded area in Figure 25) over which our sample becomes incomplete in each redshift bin, we estimate at which value of stellar mass the number of observed galaxies in our sample starts to be substantially smaller than the number predicted by each SMF.

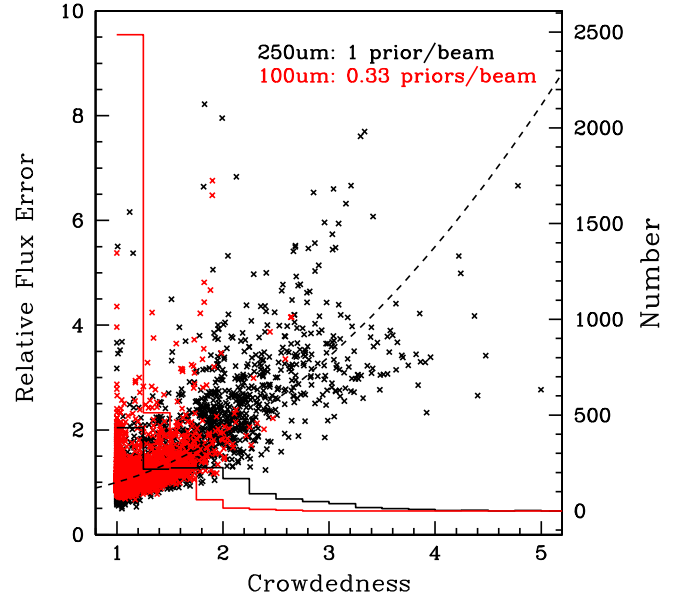
In the three bins with the highest redshifts in Figure 26, our FIR+mm sample tends to have more galaxies at the highest masses than would be predicted by the literature SMFs. In the  $3.0 < z < 4.0$  panel, the excess is not large, but in the  $4.0 < z < 5.0$  panel, our FIR+mm sample has eight more sources than is predicted by the SMFs of Grazian et al. (2015) and Song et al. (2016), or six more sources than predicted by the Davidzon et al. (2017) SMF. Note that in the GOODS-N field, there is a well-known  $z = 4.055$  proto-cluster, which includes three FIR+mm sources detected in our catalogs: GN20, GN20.2a, and GN20.2b (Daddi et al. 2009b; Tan et al. 2014). Thus small number statistics is already an important factor at these high redshifts.

Moreover, Grazian et al. (2015) discussed the importance of cosmic variance (i.e., clustering of galaxies) between the GOODS-S and UDS fields in their Figure 3 and Section 4.2.4. The SMF derived from UDS sources is much higher than the SMF derived from GOODS-S sources at the high-mass end; the UDS SMF agrees better with the SMFs of Ilbert et al. (2013) and Muzzin et al. (2013) at similar redshifts. Hence cosmic variance may also be responsible for the excess in the highest redshift bins found in this work.

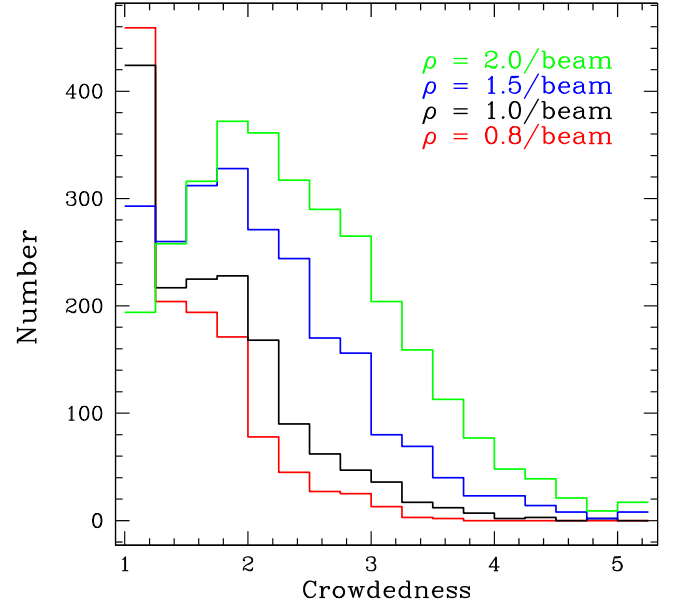
Our forthcoming work in the 2 square degree COSMOS field using the same approach (S. Jin et al. 2017, in preparation) will provide results that are less affected by cosmic variance in the high-mass and high redshift bins.



**Figure 20.** SEDs of three sources highlighted in Figure 19, to help explain the discrepancy between our photometry and that of Elbaz et al. (2011). In the panels for ID 14896 and ID 15027 panels, we show red data points from Elbaz et al. (2011) for comparison. ID 14867 has no counterpart in Elbaz et al. (2011). The best-fit redshift and reduced  $\chi^2$  are shown in the upper-left corner of each panel. When the source has a spectroscopic redshift, we fix the redshift to that value, and no error is quoted. ID 14867 and 15027 both have spectroscopic redshifts, and our deblended photometry for the latter appears to agree better with the model SEDs than that of Elbaz et al. (2011). More details of the SED fitting are given in Section 3.1. See also the caption to Figure 9.



**Figure 21.** Relative flux error as a function of *crowdedness* (normalized for low values of *crowdedness*), with values indicated on the left-Y axis. The results at 100  $\mu\text{m}$  are shown in red, while those at 250  $\mu\text{m}$  are in black. The dashed line is a reasonable fit to both data sets, scaling as  $(\text{crowdedness})^{0.5}$ . Histograms of *crowdedness* values, with values shown on the right-Y axis. At 250  $\mu\text{m}$  we are fitting a smaller number of priors that are in more crowded regions on average.

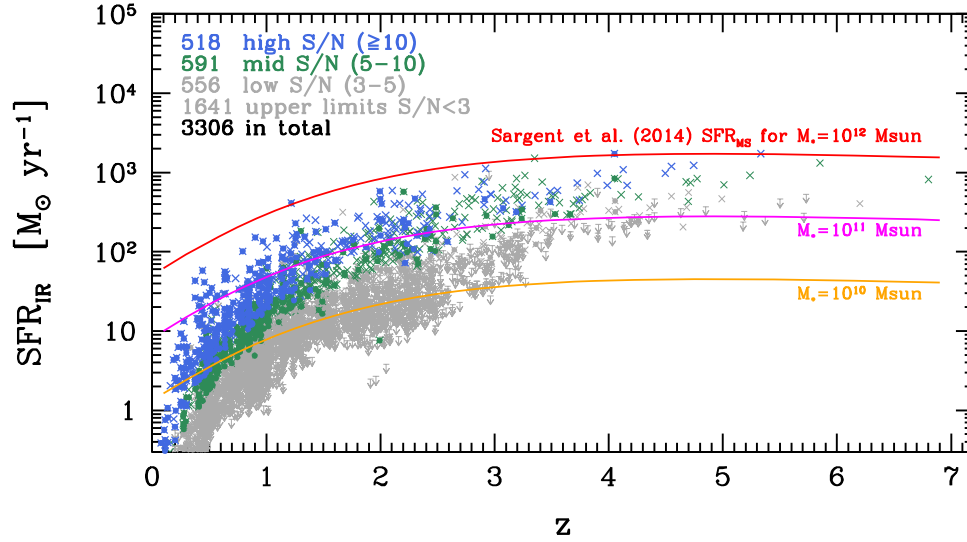


**Figure 22.** Histograms of the number of priors vs. *crowdedness* for different choices of prior source density  $\rho_{\text{beam}}$  from 0.8 to 2.0, computed here for 250  $\mu\text{m}$  and varying the choice of  $S_{\text{cut}}$ .

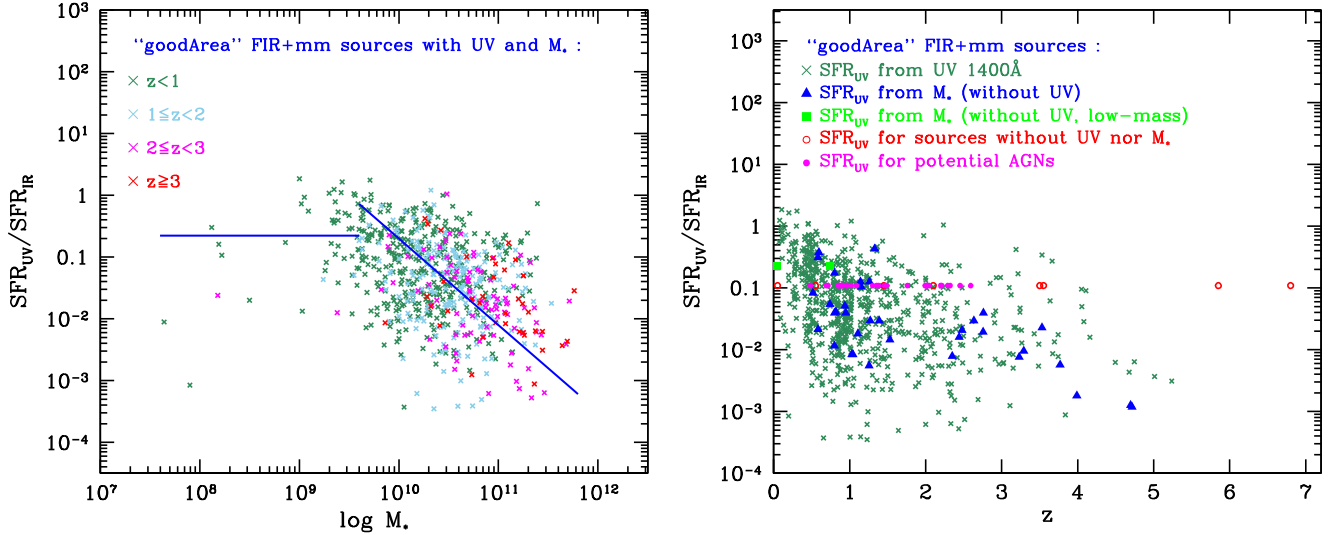
### 8.5. Comparing with SFR Histograms

In Figure 27, we analyze the SFR contribution of our sample of galaxies in bins of stellar mass. We divide our sources into several stellar mass bins and then compute the sum of SFRs in each bin.





**Figure 23.** Star formation rate (SFR) vs. redshift for all 3306 sources in the 24+radio catalog (Section 2.4). SFRs are computed from the integrated 8–1000  $\mu\text{m}$  infrared luminosities derived from their FIR+mm SEDs,  $\text{SFR} = L_{\text{IR}}/1 \times 10^{10} M_{\odot} \text{ yr}^{-1}$ , assuming a Chabrier IMF (Chabrier 2003). Colors indicate the combined S/N over the FIR+mm bands (see Equation (2)). We highlight sources with spectroscopic redshifts with dots at the center of the cross symbols. We show the empirical tracks of the MS galaxy SFR as a function of redshift at three representative MS stellar masses:  $M_* = 10^{12} M_{\odot}$ ,  $10^{11} M_{\odot}$ , and  $10^{10} M_{\odot}$ , according to Sargent et al. (2014), who compiled literature data and derived empirical MS correlations (i.e., SFR as a function of  $M_*$  and  $z$ ).

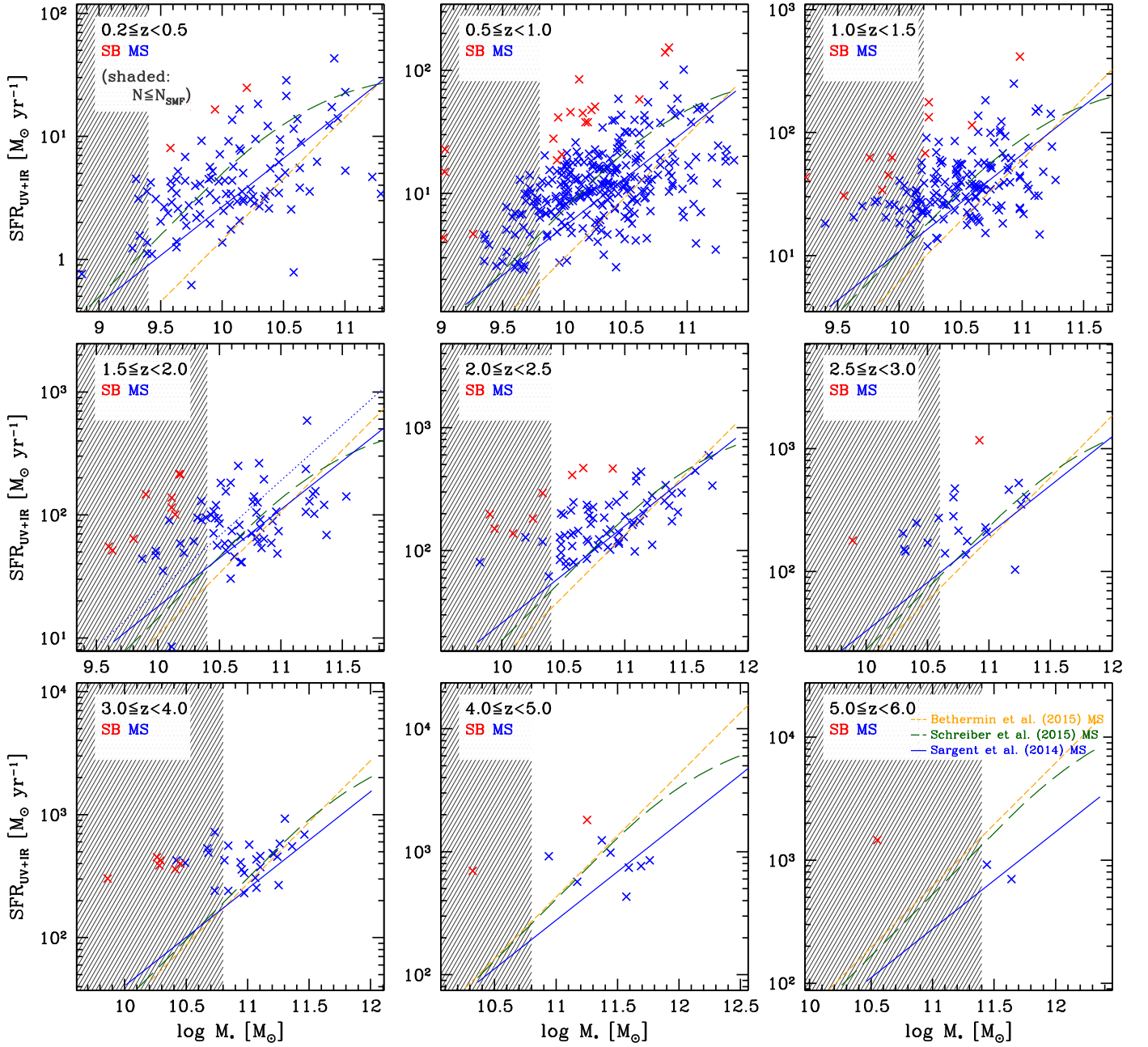


**Figure 24.** Ratio of UV-uncorrected SFR ( $\text{SFR}_{\text{UV}}$ ) and dust-obscured SFR ( $\text{SFR}_{\text{IR}}$ ) vs. stellar masses and redshift, for our FIR+mm  $\text{S/N} \geq 5$  galaxies within the *goodArea*. In the left panel, data points indicate sources for which both UV continuum fluxes and stellar masses are available, and are color-coded by redshift. In the right panel, data points are all FIR+mm sources in the *goodArea*, and colors indicate different  $\text{SFR}_{\text{UV}}$  estimation methods, depending on the availability of rest-frame UV continuum,  $M_*$ , and  $\text{SFR}_{\text{IR}}$  (see text in Section 8.2).  $\text{SFR}_{\text{IR}}$  are obtained from our IR-to-radio SED fitting, as described in Section 8.1.

For comparison, we use SMFs from the literature, first multiplying by the co-moving volume to obtain the absolute number of galaxies, and then multiplying by the SFR per unit mass predicted by the MS correlation to get the SFR contribution for each stellar mass bin (SMF-converted SFR). In this way, we are able to see how the SFR of this FIR+mm sample is distributed as a function of stellar mass, and to estimate what fraction of the co-moving SFR density this sample contributes to the whole population of star-forming galaxies at each redshift up to 6. Here we show the results computed assuming the MS correlation of Sargent et al. (2014), but the main results do not change substantially if the MSs of Béthermin et al. (2015) or Schreiber et al. (2015) were used instead.

The shape of these SMF-converted SFR histograms usually displays a peak around  $M_{\text{star}}^*$ . In the three lowest redshift bins, the predicted SFR distribution based on the Muzzin et al. (2013) SMF has the best agreement with our observations, but implies a higher SFR contribution from the most massive galaxies ( $M_* > 10^{11.0} M_{\odot}$ ), where we might be incomplete due to the limited volume of our survey. The SFR distribution predicted from the Davidzon et al. (2017) SMF matches our data well at the high-mass end, but is too low overall. The prediction based on the Ilbert et al. (2013) SMF falls in between the other two.

The FIR+mm sample is already accounting for most of the global SFR density at these redshifts. In the intermediate redshift bins ( $1.5 < z < 4.0$ ), the Ilbert et al. (2013) prediction



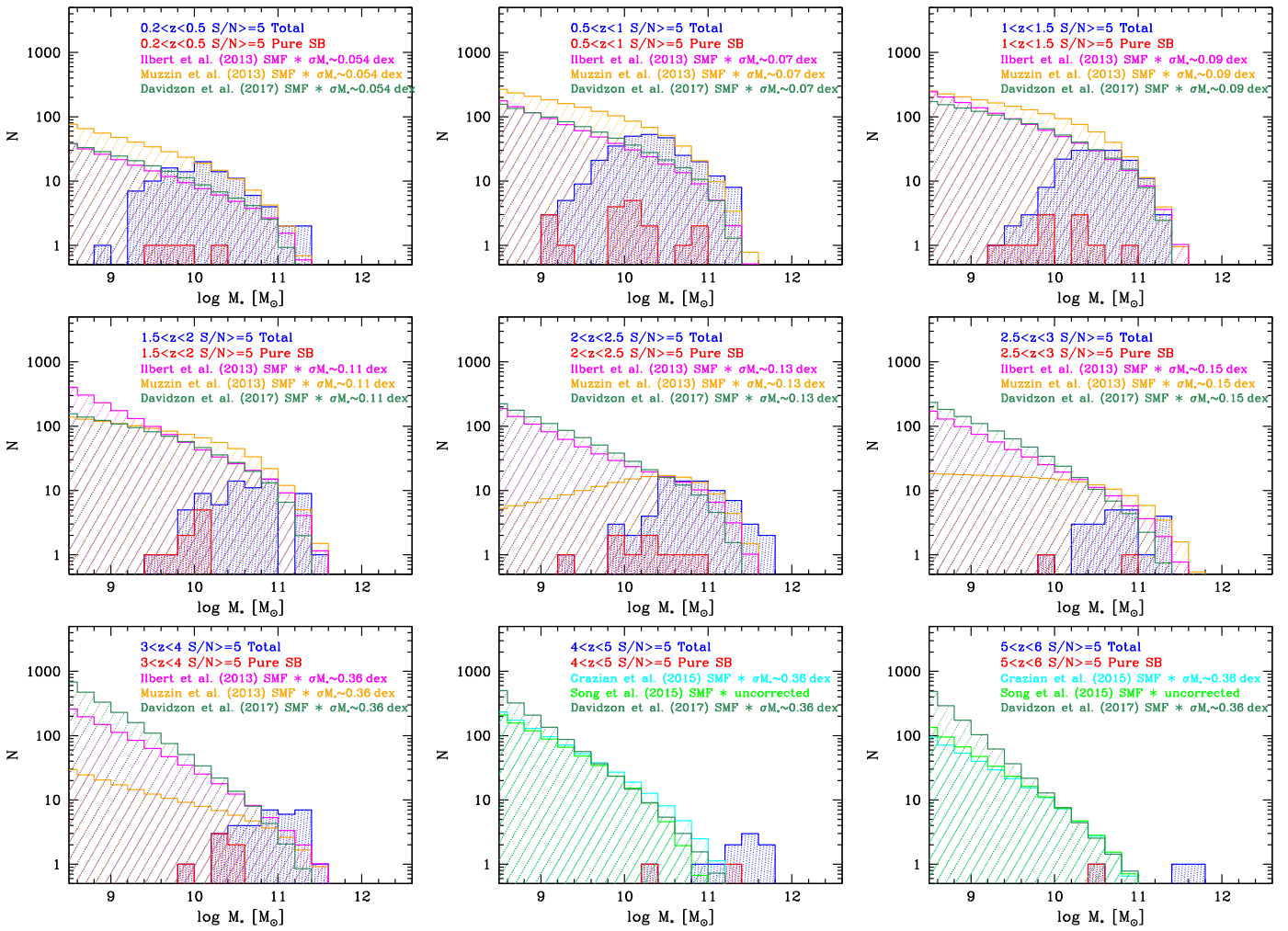
**Figure 25.** Dust-obscured plus unobscured SFR vs. stellar mass  $M_*$  for our FIR+mm sample in nine redshift bins:  $0.2 \leq z_{\text{phot}} < 0.5$ ,  $0.5 \leq z_{\text{phot}} < 1.0$ ,  $1.0 \leq z_{\text{phot}} < 1.5$ ,  $1.5 \leq z_{\text{phot}} < 2.0$ ,  $2.0 \leq z_{\text{phot}} < 2.5$ ,  $2.5 \leq z_{\text{phot}} < 3.0$ ,  $3.0 \leq z_{\text{phot}} < 4.0$ ,  $4.0 \leq z_{\text{phot}} < 5.0$ , and  $5.0 \leq z_{\text{phot}} < 6.0$ . SFR and  $M_*$  are computed assuming a Chabrier IMF (Chabrier 2003). Red symbols are sources classified as SBs during our SED fitting (Section 3.1) according to their SFR, S/N of SFR, and the distance to the MS relation of Sargent et al. (2014), while blue symbols are all other sources. The shaded area indicates where our FIR+mm sample becomes incomplete (see the evaluation of incompleteness in Section 8.4) according to our stellar mass histogram analysis in Figure 26. Three literature MSs are indicated in each panel: the blue solid line from Sargent et al. (2014), the green long-dash line from Schreiber et al. (2015), and the yellow dashed line from Béthermin et al. (2015). Note that the discrepancies between the three empirical MSs are small at lower redshifts, but not for the highest redshifts, where the difference can be as large as a factor of three.

gives the best fit to the normalization. The Muzzin et al. (2013) distribution gives a better fit to the shape at the low-mass end, which is probably due to the different stellar mass limit in Muzzin et al. (2013; e.g.,  $M_{*,95\% \text{ lim.}} = 10^{10.54} M_\odot$  at  $2 < z < 2.5$ ) compared to that in Ilbert et al. (2013; e.g.,  $M_{*,90\% \text{ lim.}} = 10^{10.01} M_\odot$  at  $2 < z < 2.5$ ), keeping in mind the current difficulties in determining the low-mass end slope of SMFs (see Appendix C of Muzzin et al. 2013). In the two highest redshift bins, the SFR distribution predicted from the Davidzon et al. (2017) SMF seems to be the closest one to the

observed SFR histogram, but the SFR excess of our FIR+mm sample is still obvious.

### 8.6. Comparing with Literature Cosmic SFR Densities

At redshifts  $z > 3$ , our understanding of the cosmic SFR density is still largely limited to UV- and optical-based studies. Direct knowledge of dust-obscured star formation at such high redshift (i.e., from the FIR to mm data) mainly comes from a small number of rarest and most ultraluminous galaxies, leaving



**Figure 26.** Stellar mass histograms of FIR+mm detected sources within the 134 arcmin<sup>2</sup> *goodArea* in nine redshift bins (blue histograms). This figure illustrates how we estimate the incompleteness of our sample at each redshift. The red histograms indicate sources classified as pure starbursts (SB) in our SED fitting. The empirical stellar mass functions (SMFs) of Ilbert et al. (2013), Muzzin et al. (2013), Grazian et al. (2015), Song et al. (2016), and Davidzon et al. (2017) are shown as the magenta, yellow, cyan, light green, and dark green histograms, respectively. These SMFs are derived for star-forming galaxies; they have been expressed for a Chabrier IMF when necessary, and convolved with typical stellar mass uncertainties ( $\sigma_{M_*}$ ) as indicated by the labels in each panel, following Appendix A of Ilbert et al. (2013; except for Song et al. 2016, which is the directly observed SMF). The SMFs of Ilbert et al. (2013) and Muzzin et al. (2013) extend up to redshift 4, while that of Grazian et al. (2015), Song et al. (2016), and Davidzon et al. (2017) probe higher redshifts. Davidzon et al. (2017) does not fully cover the highest redshift bin; there we show a simple linear redshift extrapolation.

considerable uncertainty about true SFR densities at those high redshifts (Madau & Dickinson 2014). Recent attempts to directly measure dust-obscured SFR densities from FIR and submm data have been pushed to redshift  $3 < z < 4$  (e.g., by Gruppioni et al. 2013 and Burgarella et al. 2013), and to redshift 6 (e.g., by Rowan-Robinson et al. 2016 and Bourne et al. 2017).

In this and the next section, we first directly compute the inferred SFR densities from our FIR+mm *goodArea* sample up to redshift 6 (Figure 28) and compare with results from the literature, and then estimate the incompleteness corrections using literature SMFs (Figure 29). Finally, we repeat the comparison after correcting for incompleteness in our sample (Figure 30).

We use the non-parametric  $1/V_{\max}$  method (Schmidt 1968) to estimate the SFR density of our FIR+mm sample in each redshift bin in order to eliminate the Malmquist bias. For each source, we compute the farthest redshift  $z_{\max}$  at which it can still be detected in our catalog (i.e.,  $S/N_{\text{FIR+mm}} \geq 5$ ). Then for each redshift bin (lower and higher boundaries are  $z_1$  and  $z_2$ ), we compute the co-moving volume for each source ( $V_{\max}$ ) from  $z_1$  to  $\min(z_{\max}, z_2)$

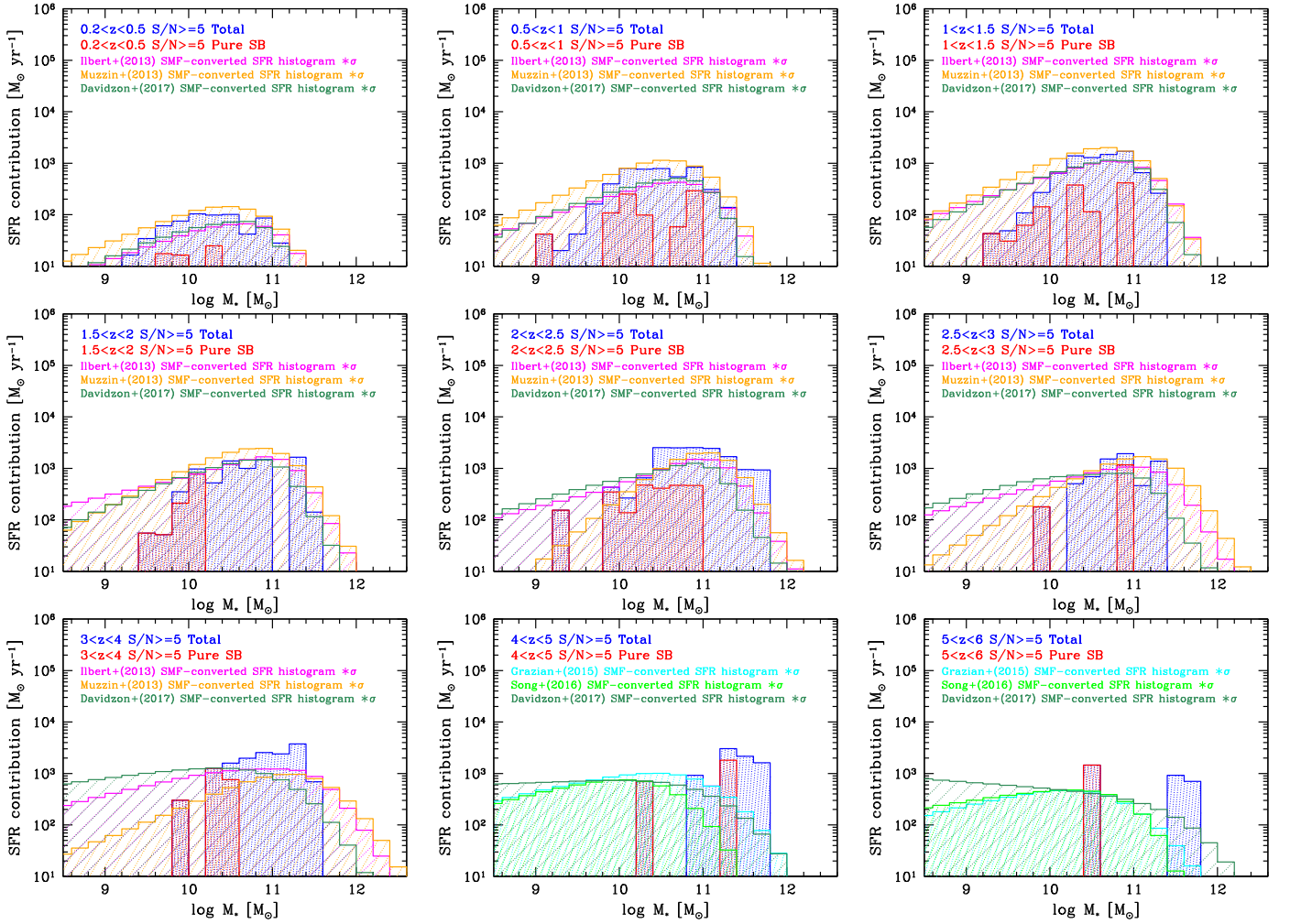
(i.e., the smaller value; e.g., Pozzetti et al. 2003). Then the SFR density of this sample is the sum of all individual  $\text{SFR}/V_{\max}$  in each redshift bin. To account for more realistic uncertainties, we compute  $\text{SFR}/V_{\max}$  in a bootstrapping method: we randomly select the same number of sources in each redshift bin, allowing duplication, and then repeat 1000 times to compute the mean and rms (i.e., the uncertainty of the measured SFR density).

Figure 28 shows the SFR densities directly measured from our observed sample, and compares them to previous results from the literature (Gladders et al. 2013; Madau & Dickinson 2014). The fluctuations from bin to bin at  $0 < z < 3$  (where we expect a smooth redshift dependence) are significant at the  $\sim 2\sigma$  level and suggest total uncertainties in our determinations of order of  $\sim 0.2$  dex, probably dominated by cosmic variance (although note that the bins do not have equal co-moving volumes).

### 8.7. Incompleteness Corrections for Cosmic SFR Densities

We estimate the completeness of our measured SFR density from the FIR+mm sample to the total cosmic SFR density in





**Figure 27.** As in Figure 26, but showing the measured SFR contribution from each mass bin. We convert the SMF histograms into SFR histograms by multiplying by the specific-SFR defined by MS of Sargent et al. (2014) (see Section 8.6). This figure shows the direct comparison between the data and the empirical expectations from the SMFs without any renormalization. The renormalized comparison, used to assess completeness correction, can be seen in Figure 53.

each redshift bin in the following way, with the help of the SFR histograms in bins of stellar masses. We assume that SMF can predict the total cosmic SFR density based on the assumption of MS correlation (e.g., Sargent et al. 2014); thus the SMF-converted SFR histogram should be able to fit the observed SFR histogram when the FIR+mm sample is complete at certain stellar mass bin (e.g., as indicated by Figure 27). The fitting is done by adjusting the normalization of SMF-converted SFR histogram (see Figure 53 in the Appendix), so as to achieve a minimum  $\chi^2$  in the stellar mass range where the FIR+mm sample is at least 50%-complete in stellar mass. The 50%-complete mass ( $M_{*,50\% \text{ lim.}}$ ) is determined by the renormalized version of Figure 26 (Figure 53 in the Appendix): we calculate above which stellar mass the number of FIR+mm galaxies starts to be larger than half of what is predicted by SMF (i.e.,  $N_{\text{FIR+mm}} \geq 0.5 \times N_{\text{SMF}}$ ).

We then integrate the renormalized SMF-converted SFR histogram down to  $10^8 M_\odot$  (as was done in Ilbert et al. 2013 and Muzzin et al. 2013) to obtain an estimate of the total cosmic SFR density. The ratio of the total observed SFR to the total cosmic SFR gives the completeness fraction ( $f_{\text{comp.}}$ ).

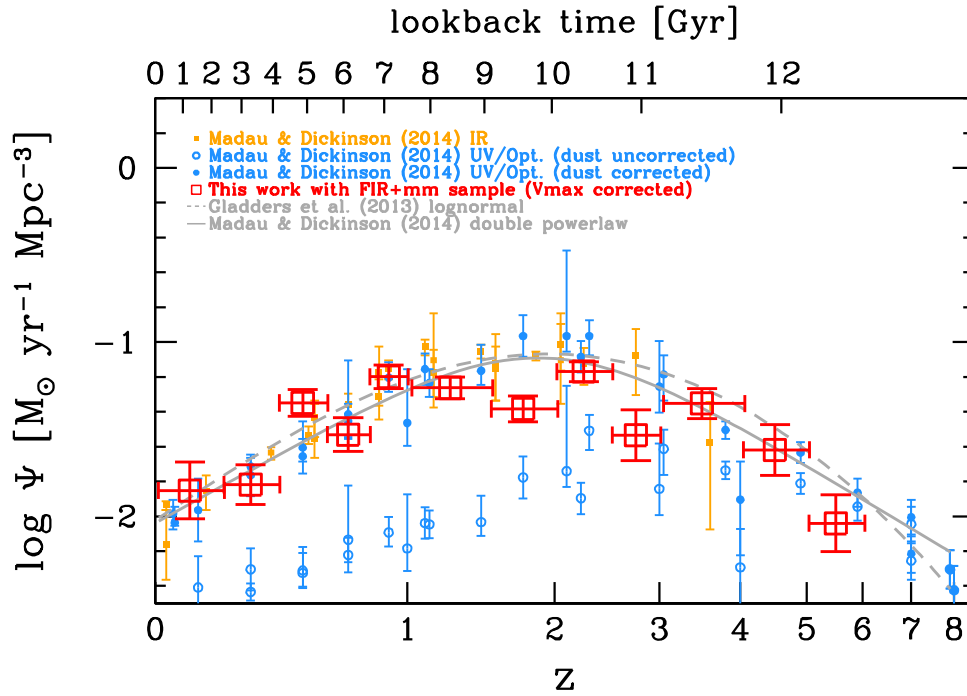
Note that, as mentioned at the end of Section 8.2, the directly observed SFRs for our FIR+mm sample are  $\text{SFR}_{\text{UV+IR}}$ . We emphasize that the aim of the incompleteness correction is to

account for the  $\text{SFR}_{\text{UV+IR}}$  from fainter/lower-mass galaxies that are not included in our sample. Although one might consider correcting the observed IR-only SFRs in the same fashion, we caution that this may not be appropriate if fainter/lower-mass galaxies have less dust attenuation, as is observed at  $z < 3$  (e.g., Pannella et al. 2015).

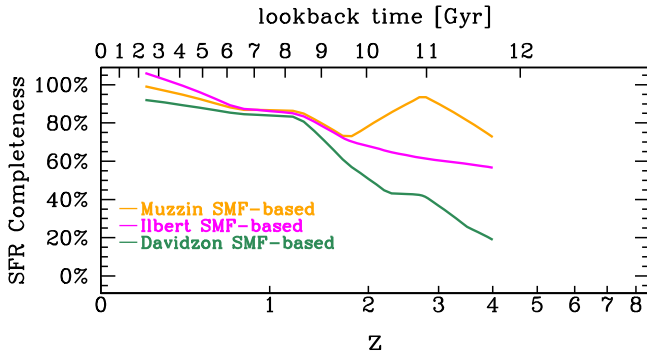
In Figure 29, we present the cosmic SFR completeness fractions versus redshift, estimated using three different SMFs from the literature. We present the incompleteness-corrected total SFR densities in Figure 30.

At  $z < 2$ , the completeness of our sample drops linearly with  $\log(1+z)$ , and the three SMF-based results are fully consistent. However, at  $z > 2$ , the three SMF-based estimates differ by large factors, largely due to the different low-mass slopes of the literature SMFs. The SMF from Muzzin et al. (2013) has a shallower slope at low masses, and fits our observed histograms much better, leading to a completeness factor of about 80% at  $z \sim 3$ . The SMFs from Ilbert et al. (2013) and Davidzon et al. (2017) have steeper low-mass slopes, and hence lead to much larger incompleteness corrections (see Table 3).

At  $z > 4$ , the completeness fractions that we estimate from the renormalized SMFs are very low (see Appendix, Figure 53), ranging from  $\sim 0.2\%$  to  $\sim 5\%$ . If we do not re-normalize the SMFs but just integrate the SMF-based SFR histogram (i.e., as shown in



**Figure 28.** Co-moving SFR densities ( $\Psi$ ) of our FIR+mm sample (red symbols) compared with values from the literature. The horizontal error bars represent the redshift bin ranges, and vertical error bars indicate the uncertainties. The measurement details (i.e.,  $1/V_{\max}$  and bootstrapping methods) are discussed in Section 8.6. The yellow solid squares are the IR-based measurements from Madau & Dickinson (2014). Blue filled circles are the rest-frame UV-based dust attenuation-corrected measurements, while the blue open circles are rest-frame UV measurements with no dust attenuation correction. The dashed gray curve shows the best-fit log-normal function of Gladders et al. (2013), while the solid gray curve is the best-fit double power-law function given in Madau & Dickinson (2014).



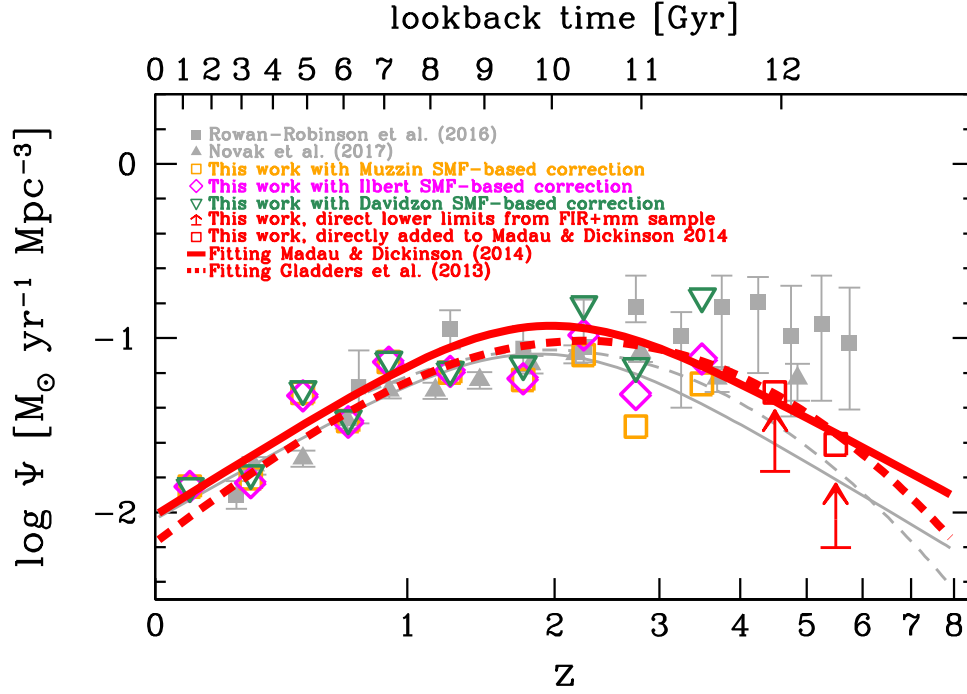
**Figure 29.** Completeness in SFR for our FIR+mm sample, estimated based on fitting/re-normalizing the SMFs. See details in section 8.7 and the Appendix, Figure 53.

Figure 27), we would get a total cosmic SFR density that is fully consistent with the fit by Madau & Dickinson (2014) at  $z \sim 5$ , and would not reflect the excess SFR that we see in the observed SFR histograms. The main reason is that at  $z > 4$  our FIR+mm sample is almost entirely disjoint from the UV-based samples that have traditionally been used to estimate the cosmic SFRD, as we will demonstrate in the next sections. Thus to estimate the total cosmic SFR density at this high redshift range, we need to add our directly observed FIR+mm SFR density to the UV-based SFR density. In Figure 30, we show the summed total cosmic SFR densities at  $z > 4$  as the red open squares. Our directly observed FIR+mm SFR densities are also shown as the red arrows for comparison. We list all corrected cosmic SFR density values in Table 3.

Our corrected measurements imply total SFR densities that are consistent with those reported in Madau & Dickinson (2014) at  $0.5 < z < 2$ , but possibly up to 0.4 dex higher at  $2 < z < 4$ , where the results are less reliable and depend strongly on the assumed SMF. However, we also note that several recent FIR,

(sub)mm and radio studies also seem to find similar trends. In Figure 30, we show the recent cosmic SFR density measurements from Rowan-Robinson et al. (2016) and Novak et al. (2017) for comparison. Based on much wider (20 degree<sup>2</sup>) but shallower FIR surveys, Rowan-Robinson et al. (2016) studied the FIR luminosity function with 3035 SPIRE 500  $\mu\text{m}$  sources. Their cosmic SFR densities (the gray filled squares) are much higher than the UV estimates at  $3.5 < z < 6$  (i.e., the Madau & Dickinson 2014 curve) by factors of  $\sim 2\text{--}6\times$ . They assumed a fixed 500  $\mu\text{m}$  luminosity function at  $z > 3.5$  (Saunders et al. 1990, functional form with fixed parameters  $\alpha = 1.2$  and  $\sigma = 0.60$ ), but whether such fixed form is applicable is still controversial. Novak et al. (2017) used deep VLA imaging data in the 2 deg<sup>2</sup> COSMOS field to derive and integrate the 3 GHz luminosity function (assuming pure luminosity evolution) up to  $z = 5\text{--}6$ , and estimated the cosmic SFR density assuming a redshift-dependent FIR-radio correlation. Their results are shown with gray triangles in Figure 30, and are in good agreement with the Madau & Dickinson (2014) curve and the average of our incompleteness-corrected data points at  $z < 4$ . However, at  $z \sim 5$ , their cosmic SFR density is  $\sim 3\times$  the value from Madau & Dickinson (2014).

Recently, Bourne et al. (2017) and Dunlop et al. (2017) have investigated dust-obscured SFR densities by stacking submm and mm data for faint galaxies selected from deep *HST* imaging data. Bourne et al. (2017) used SCUBA2 data at 850  $\mu\text{m}$  and 450  $\mu\text{m}$  covering 230 arcmin<sup>2</sup> in three of the CANDELS fields, stacking emission for galaxies selected from the 3D-*HST* catalog of Momcheva et al. (2016). Dunlop et al. (2017) used ALMA 1.3 mm data covering 4.5 arcmin<sup>2</sup> in the HUDF, stacking for galaxies selected from the deep HUDF *HST* data. They convert the stacked single (sub)mm band flux into dust-obscured SFRs using submillimeter galaxy template SEDs, and then sum this with the raw UV SFRs, uncorrected for dust



**Figure 30.** Incompleteness-corrected cosmic SFR densities based on our FIR+mm sample. We estimate the incompleteness using several different SMFs from the literature (Ilbert et al. 2013; Muzzin et al. 2013; Grazian et al. 2015; Song et al. 2016; Davidzon et al. 2017) and assume the MS correlation fitted by Sargent et al. (2014). At  $z < 4$ , we show the incompleteness-corrected cosmic SFR density derived from each SMF-based correction using different colors and symbols (see labels in the figure). At  $z > 4$ , we show lower limits as well as the summed total cosmic SFR density that combines our FIR+mm sample and the UV-based samples compiled by Madau & Dickinson (2014). See details in Section 8.7. The two gray curves are the same as in Figure 28, but here we also show two red thick curves (solid curve—double power-law function as in Madau & Dickinson 2014; dashed curve—log-normal function as in Gladders et al. 2013), which are our new best-fit to all the colored (i.e., green, magenta, yellow and red) data points (details in the text). The gray data points show other results from the recent literature derived from far-IR or radio data extending out to  $z = 5$ –6 (Rowan-Robinson et al. 2016; Novak et al. 2017). All values assume a Chabrier IMF (or are converted to Chabrier IMF when needed).

attenuation. At  $z < 2$ , the total cosmic SFR densities that they derive agree well with the compilation of Madau & Dickinson (2014), similar to our own findings here. At higher redshift, their results are  $\sim 0.1$ – $0.2$  dex larger than the Madau & Dickinson (2014) curve, but still  $\sim 0.2$ – $0.3$  dex lower than the present work. It is possible that the most luminous IR-bright galaxies at  $z > 2$ – $3$  were not part of their stacked samples, which were selected based on near-IR ( $1.6 \mu\text{m}$ ) *HST* data.

As a further step, we make a simple parametric fit to our incompleteness-corrected data points (i.e., all colored data points in Figure 30) using the Madau & Dickinson (2014) double power-law and Gladders et al. (2013; see also Abramson et al. 2016) log-normal equations. From the  $\chi^2$  distribution statistics, we compute the median value of each parameter within the  $\chi^2_{\min}$  to  $\chi^2_{\min} + d\chi^2$  range,<sup>34</sup> and then obtain the following newly fitted equations and parameters:

$$\Psi = \Psi_0 \frac{(1+z)^{3.0}}{1 + [(1+z)/2.9]^{5.6}} M_{\odot} \text{ yr}^{-1} \text{ Mpc}^{-3}$$

$$\Psi_0 = 0.00587 \text{ (Chabrier IMF)} \quad (3)$$

(the red thick solid curve in Figure 30), and

$$\Psi = \frac{A_0}{t\sqrt{2\pi\tau^2}} \exp\left[-\frac{(\ln(t) - T_0)^2}{2\tau^2}\right] M_{\odot} \text{ yr}^{-1} \text{ Mpc}^{-3}$$

$$t \text{ is the cosmic age in Gyr, } T_0 = 1.50, \tau = 0.66$$

$$A_0 = 0.575 \text{ (Chabrier IMF)} \quad (4)$$

(the red thick dashed curve in Figure 30).

The new curves are similar to the previous fits at  $z < 1$ . The double power-law (solid curve) is higher at  $z \sim 2$ , but the log-normal (dashed curve) peaks at values similar to the original fits at  $z \sim 2$ , where our SMF-converted SFR densities agree well with the literature data. At  $z > 3$ , both new curves can generally fit the new FIR/radio-based data points. But they diverge in the highest redshift bin, where Rowan-Robinson et al. (2016) measured a very high value, but we derive a lower value of the FIR SFR density. There might still be some less extreme dusty galaxies, with lower infrared luminosities and SFRs, that are missed in both our FIR+mm sample and in UV-based samples. Such objects might be recovered by future ALMA/NOEMA surveys.

#### 8.8. Uncertainties in the Cosmic SFR Density Studies

Using our new FIR+mm photometric catalog, the completeness-corrected cosmic SFR densities are quite well constrained up to redshifts  $z \sim 2$ – $3$ . Variations up and down around the average trend reflect small sample statistics together with cosmic variance in the relatively small GOODS-North field.

At  $z > 3$ , despite our concerted effort on photometry, the uncertainties are still much larger, not only due to cosmic variance but also due to other factors that we briefly discuss here. There are remaining limitations in the photometry, even for bright sources due to residual blending. For example, we have evaluated the incompleteness in recovering IR sources as a function of their SFR/masses and redshifts. Figures 26 and 27 show that this is not likely to be a big issue at  $z < 3$  for bright galaxies (while we correct for faint galaxies), but this clearly becomes more important at  $z > 3$  (Figure 29). This is further exemplified by

<sup>34</sup>  $d\chi^2 = 3.53$  for three-parameter fitting, and  $d\chi^2 = 4.72$  for four-parameter fitting (e.g., Press et al. 1992, chapter 15.6).



**Table 3**  
GOODS-N FIR+mm Star Formation Rate Densities

$z_{\text{low}}$	$z_{\text{high}}$	$\Psi_{\text{SFR,obs,UV+IR}}$	$\Psi_{\text{SFR,obs,IR-only}}$	$\sigma_{\Psi_{\text{SFR,obs}}}$	$\Psi_{\text{SFR,corr.}}$	SMF <sub>corr.</sub>	$f_{\text{comp.}}$
0.0	0.2	−1.852	−2.00	0.164	−1.852	Muzzin	100.0%
					−1.852	Ilbert	100.0%
					−1.852	Davidzon	100.0%
0.2	0.4	−1.818	−1.93	0.114	−1.805	Muzzin	97.0%
					−1.830	Ilbert	102.8%
					−1.776	Davidzon	90.7%
0.4	0.6	−1.349	−1.42	0.077	−1.315	Muzzin	92.5%
					−1.331	Ilbert	96.0%
					−1.293	Davidzon	88.0%
0.6	0.8	−1.531	−1.56	0.096	−1.476	Muzzin	88.1%
					−1.482	Ilbert	89.3%
					−1.462	Davidzon	85.4%
0.8	1.0	−1.199	−1.22	0.067	−1.137	Muzzin	86.8%
					−1.137	Ilbert	86.8%
					−1.125	Davidzon	84.2%
1.0	1.5	−1.263	−1.30	0.062	−1.199	Muzzin	86.4%
					−1.191	Ilbert	84.8%
					−1.183	Davidzon	83.2%
1.5	2.0	−1.383	−1.41	0.075	−1.239	Muzzin	71.8%
					−1.233	Ilbert	70.8%
					−1.150	Davidzon	58.5%
2.0	2.5	−1.170	−1.19	0.057	−1.098	Muzzin	84.8%
					−0.983	Ilbert	65.0%
					−0.806	Davidzon	43.3%
2.5	3.0	−1.534	−1.54	0.146	−1.509	Muzzin	94.4%
					−1.322	Ilbert	61.5%
					−1.161	Davidzon	42.3%
3.0	4.0	−1.353	−1.38	0.086	−1.263	Muzzin	81.3%
					−1.120	Ilbert	58.6%
					−0.762	Davidzon	25.7%
4.0	5.0	−1.619	−1.64	0.145	−1.313	MD14	~49.4%
5.0	6.0	−2.040	−2.07	0.164	−1.610	MD14	~37.1%

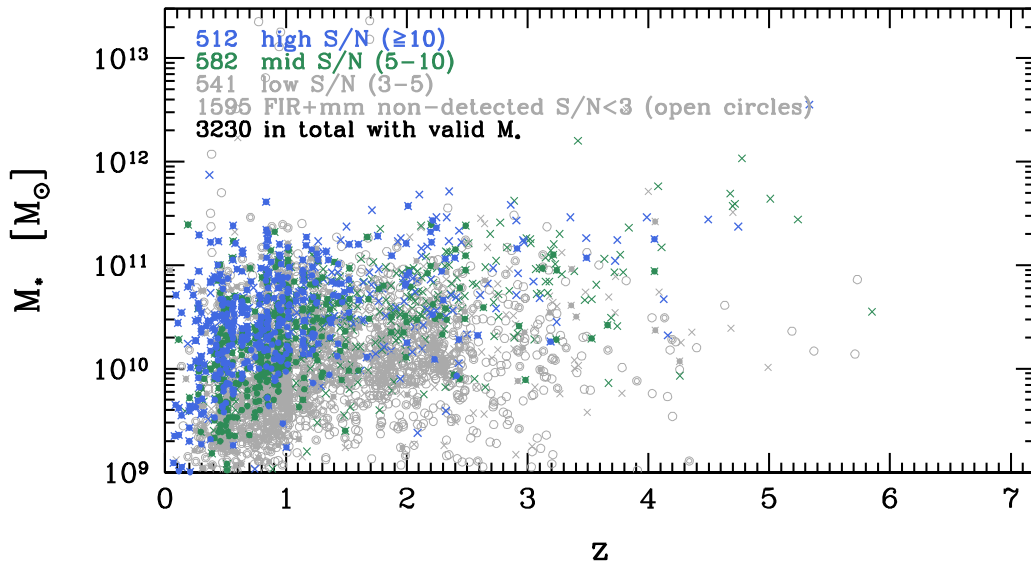
**Note.** The measured SFR density ( $\Psi_{\text{SFR,obs,UV+IR}}$  and  $\Psi_{\text{SFR,obs,IR-only}}$ , in units of  $M_{\odot} \text{ yr}^{-1} \text{ Mpc}^{-3}$ ) and uncertainty ( $\sigma_{\Psi_{\text{SFR,obs}}}$ ) for the observed FIR+mm galaxies in this work as described in Section 8.6, as well as the incompleteness-corrected cosmic SFR density ( $\Psi_{\text{SFR,corr.}}$ ) and the completeness percentage in SFR ( $f_{\text{comp.}}$ , i.e.,  $=10^{(\Psi_{\text{SFR,obs,UV+IR}} - \Psi_{\text{SFR,corr.}}) \times 100\%}$ ) as described in Section 8.7. We use literature SMFs and the Sargent et al. (2014) MS correlation to estimate the completeness percentage; therefore each SMF leads to a corrected value of  $\Psi_{\text{SFR,corr.}}$  at each redshift bin. The SMF<sub>corr.</sub> column indicates which SMF is used (Ilbert et al. 2013; Muzzin et al. 2013; Grazian et al. 2015; Song et al. 2016; Davidzon et al. 2017), except for the last two rows at  $z > 4$ , where it becomes prohibitive to estimate any incompleteness correction. Given the evidence that our FIR+mm sample does not substantially overlap with the UV-based LBG samples at  $z > 4$  (see Section 8.9), we correct the  $z > 4$  cosmic SFR densities by adding the observed SFR densities from our FIR+mm sample to the UV-based, attenuation-corrected SFR densities from LBG samples (i.e., the values reported in Madau & Dickinson 2014, MD14). The percentages ~49.4% and ~37.1% thus represent the SFR ratio of our sample to the sum of our sample and the LBGs (Madau & Dickinson 2014, MD14).

the “additional sources” that we found and neglected in the SFRD study (Section 3.4). Assuming that some of these are not spurious, some of them are likely to have redshifts  $z > 3$ , as shown by the example of HDF850.1, which is not included in the SFRD calculation. The limited sensitivity of the IR data also is a factor, permitting direct detections of sources only with  $\text{SFR} > 500 M_{\odot} \text{ yr}^{-1}$  at  $z > 4$ . This may be far brighter than the threshold at which UV selection starts to include the large majority of galaxies, although the flux range for which this is true is currently unknown at  $z > 3-4$ . In any case, this may be another source of significant incompleteness, even for UV+IR samples at these redshifts. Meanwhile the accuracy of photometric redshifts can also be a problem. The photometric data used to estimate redshifts are very good for well-studied deep survey fields like GOODS-North, but catastrophic photometric redshift failures might still occur (e.g., in the case of red galaxies with apparent photometric redshifts,  $z > 3-6$ ). It is noteworthy that

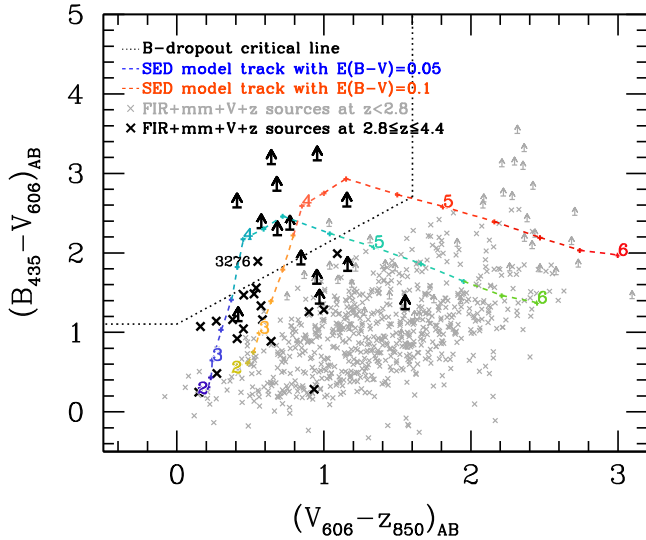
some of the most distant and luminous star-forming galaxy candidates that we find, with  $z > 5$  (see next sections), are also embarrassingly massive (see Figure 31). If their redshifts are correct, which must be tested with spectroscopy, these objects could pose interesting challenges for galaxy formation models, or may demonstrate limitations in our ability to derive accurate stellar masses for such objects (see Section 8.10 for more discussions about these sources).

### 8.9. Comparing with Lyman-break Galaxies (LBGs)

We have found a substantial population of IR-detected galaxies all the way out to redshifts  $z = 3-6$ , and have estimated that their contribution to the SFR density of the universe is significant and comparable to or larger than that derived from samples selected at UV rest-frame wavelengths. It is therefore interesting to discuss the possible overlap between



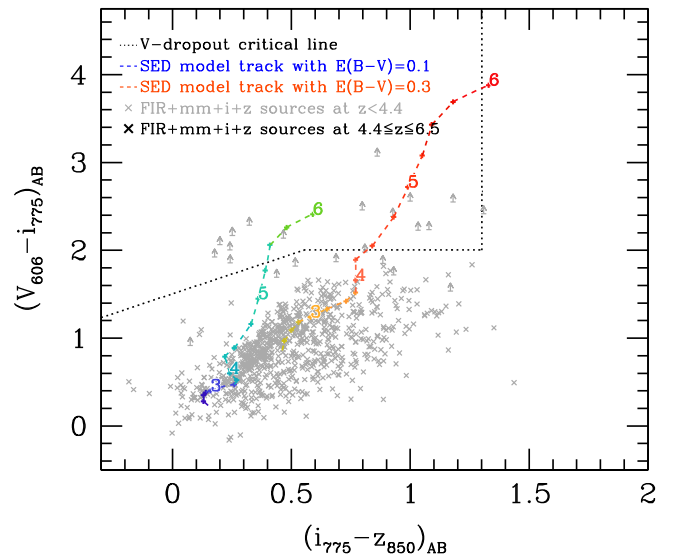
**Figure 31.** Similar to Figure 23 but showing stellar masses vs. redshift for objects in our prior list. Only sources having stellar mass estimates from optical/near-IR catalogs are shown: the number of plotted sources is thus smaller here than in Figure 23. The colors and symbol shapes coding is the same as in Figure 23 and reflects the FIR+mm properties, not stellar masses (i.e., upper limits mean that the object was not detected at FIR+mm, not that  $M^*$  is an upper limit, etc.).



**Figure 32.** Color-color diagram for identifying  $B$ -dropouts at  $2.8 \leq z \leq 4.4$  (Vanzella et al. 2009; Dahlen et al. 2010). Data points are FIR+mm detected galaxies which are also detected at  $V_{606}$  and  $z_{850}$  bands. Gray points indicate galaxies with  $z < 2.8$ , and black points indicate galaxies at  $2.8 \leq z \leq 4.4$  (e.g., Figure 1 of Vanzella et al. 2009). Galaxies not detected in  $B_{435}$  are shown as upward arrows ( $1\sigma$  lower limits). The dotted line indicates the  $B$ -dropout criteria (Equation (5)). Two colored curves indicate two model tracks calculated from our SED templates (Figure 2), with  $E(B-V) = 0.05$  and  $0.1$  as two examples. The small colored number at each node of the track indicates the redshift.

galaxy samples selected at IR and UV wavelengths in order to evaluate whether they trace *independent* components of the cosmic SFR density (which should then be summed to derive the total), or if we are measuring the same population.

At  $z > 3$ , the 912 Å Lyman limit and 1216 Å Lyman  $\alpha$  forest spectral breaks are redshifted through optical passbands, causing the galaxies to become very faint and to “drop out” at bluer wavelengths while still being clearly detected in redder filters. Optical color-color diagrams can then be used to select samples of LBGs in specific redshift ranges based on these “dropout” color signatures. For the *HST* ACS filters that were used to observe GOODS-N, most  $B$ -dropout LBGs fall into the



**Figure 33.** Similar to Figure 32 but for  $V$ -dropouts at  $4.4 \leq z \leq 6.5$  (Vanzella et al. 2009). No boldface point is seen here, because none of the FIR+mm galaxies in our sample at  $4.4 < z < 6.5$  are detected by *HST* in these optical bands.

redshift range  $2.8 \leq z \leq 4.4$ , while  $V$ -dropout LBGs are found to have  $z \approx 4.4$ – $6.5$  (Giavalisco et al. 2004; Vanzella et al. 2009; Dahlen et al. 2010). However, the Lyman-break technique relies on detecting and measuring rest-frame UV light, and may therefore miss a fraction of dusty galaxies whose UV emission is strongly attenuated and reddened. To investigate how dusty galaxies are missed in LBG samples, we compare our GOODS-N FIR+mm detected sources at  $z = 2.8$ – $4.4$  and  $z = 4.4$ – $5.5$  with the  $B$ - and  $V$ -dropout criteria (e.g., Dickinson et al. 2004; Giavalisco et al. 2004; Stark et al. 2009; Vanzella et al. 2009; Bouwens et al. 2012).

We use the *HST* optical to near-infrared photometry from Skelton et al. (2014; cross-matched in Section 2) and use the color-color criteria from Vanzella et al. (2009) to select  $B$ - and  $V$ -dropout LBGs in GOODS-N:

1. *B*-dropout

$$\begin{cases} (B_{435} - V_{606}) > (V_{606} - z_{850}) + 1.1 \\ \text{and } (B_{435} - V_{606}) \geq 1.1 \\ \text{and } (V_{606} - z_{850}) \leq 1.6 \end{cases} \quad (5)$$

2. *V*-dropout

$$\begin{cases} (V_{606} - i_{775}) > 0.9 \times (i_{775} - z_{850}) + 1.5 \\ \text{or } (V_{606} - i_{775}) > 2.0 \\ \text{and } (V_{606} - i_{775}) > 1.2 \\ \text{and } (i_{775} - z_{850}) < 1.3 \\ \text{and } S/N(B_{435}) < 2 \\ \text{and } S/N(i_{775}) \geq 3. \end{cases} \quad (6)$$

In Figure 32, FIR+mm detected sources at any redshift that also have  $V_{606}$  and  $z_{850}$  three-sigma detections are plotted in the  $B_{435} - V_{606}$  versus  $V_{606} - z_{850}$  diagram. The subscripts indicate the *HST* observing bands (i.e., F435W for *B* band, F606W for *V* band, and F850LP for *z* band). We also show two modeled tracks as the colored curves, which are calculated by redshifting SED templates (BC03+DL07; e.g., Figure 2). Here we show two example model tracks with  $E(B - V) = 0.05$  and 0.1. Heavier dust attenuation will make the model track shift toward redder colors. Data points within the upper-left region identified by the dotted lines indicate galaxies that meet the  $B_{435}$ -dropout color criteria.

Among all the FIR+mm and  $V_{606}$  and  $z_{850}$  detected galaxies, we highlight in Figure 32 the 32 ones having  $2.8 \leq z \leq 4.4$ . Only 9 of them can be identified as *B*-dropouts (we use  $1\sigma$  limits, as in Vanzella et al. 2009). In comparison, there are a total of 78 FIR+mm galaxies with photometric redshift within  $2.8 \leq z \leq 4.4$  (59 have valid optical photometry in 3D-*HST* catalog, often thus with near-IR WFC3 detections but lacking ACS detections in the optical). Therefore, only half of our FIR+mm sample at  $2.8 < z < 4.4$  has a significant *HST* optical counterpart in *V* and *i* to start with, and only about 10% of the parent sample (20% of those with optical counterparts) could be identified as *B*-dropouts at the depths of the current *HST* imaging. The optically detected FIR+mm galaxies with  $2.8 < z < 4.4$  fail to be selected as *B*-dropouts for a number of reasons: photometric noise affecting these faint detections, effect of photometric redshifts (some objects might be slight out of the boundary, in reality), lack of depth in the optical data (upper limits in the *B* band are not always stringent enough). In addition, even for the very small number of galaxies that can be selected as *B*-dropouts, the rest-frame UV-based SFR significantly underestimates the dust-obscured SFR. For example for ID 3276, a source that is detected in *B* band and is within the *B*-dropout region in Figure 32 (with its ID number labeled in the figure), the optical/near-IR (rest-frame UV) based SFR, corrected for dust attenuation, is  $\sim 32 M_{\odot} \text{ yr}^{-1}$  (stellar mass  $10^{11.06} M_{\odot}$ ) in the 3D-*HST* catalog. For comparison, its FIR+mm based SFR is  $256 \pm 37 M_{\odot} \text{ yr}^{-1}$  in our FIR+mm SED fitting.

In Figure 33, we show the distribution of FIR+mm galaxies with  $i_{775}$  and  $z_{850}$  detections in the  $V_{606} - i_{775}$  versus  $i_{775} - z_{850}$  diagram. None of the seven galaxies in our sample with a photometric redshift within the *V*-dropout range of  $4.4 \leq z \leq 6.5$  is detected in  $i_{775}$  and  $z_{850}$ . There are some galaxies within the *V*-dropout color selection region for which the photometric redshift estimate is  $z < 4.4$ .

All in all, there is very little overlap between our FIR+mm-selected sample of high redshift galaxies and dropout samples. For those few FIR+mm-bright objects that can be selected as LBGs, their UV-estimated SFRs are largely underestimated compared to those derived from their far-infrared luminosities. We thus conclude that the SFR density contributions we have derived from IR data are largely independent from the ones estimated from UV-selected samples, at least at  $z \gtrsim 3.5$ .

8.10. A Sample of Candidate  $z \geq 5$  Galaxies

In our FIR+mm catalog, we find three sources that have FIR+mm SED photometric redshifts  $z \geq 5$ , and whose SFRs and stellar masses appear to be consistent with the MS at these redshifts (at least within the current, large uncertainties on its locus). These are ID 14914 (which is not in the *goodArea*), ID 6406, and ID 12646 (the right-most blue data point in Figure 23, our highest redshift candidate, with  $z \sim 6.8$ ). We discuss the properties of these three particularly interesting sources in some details below.

ID 14914 is detected in all FIR+mm bands. Multi-wavelength SED and cutout images are presented in Figures 34 and 35. Weak extended emission can be seen in the *HST* ACS and WFC3 images. Two sources from the 3D-*HST* catalogs are very close to ID 14914. We cross-matched the closer ( $0''.585$ ) 3D-*HST* source, to the north-east, as the counterpart. It is clearly detected by IRAC, as well as in the (sub)mm and at 20 cm. Blending is not significant. The 3D-*HST* ID number and other properties derived from our analysis are listed below:<sup>35,36</sup>

$$\begin{aligned} \text{ID}_{\text{IRAC}} &= 14914 \\ \text{ID}_{3\text{D-HST}} &= 7338 \\ z_{\text{FIR+mm}} &= 5.33 \pm 0.18 \\ z_{3\text{D-HST}} &= 5.57 \text{ (5.335--5.90)} \\ M_{*,3\text{D-HST}} &= 10^{12.55} M_{\odot} \\ M_{*} &\approx 10^{11.68} M_{\odot} \\ \text{SFR}_{3\text{D-HST}} &= 436.5 M_{\odot} \text{ yr}^{-1} \\ \text{SFR}_{\text{FIR+mm}} &= 1724 \pm 132.1 M_{\odot} \text{ yr}^{-1} \\ \langle U \rangle &\approx 45. \end{aligned} \quad (7)$$

This source is also identified as AzGN04 and GN1200.12 in the 1.1–1.2 mm catalogs of Perera et al. (2008), Greve et al. (2008), and Penner et al. (2011). We show the photometry of Chapin et al. (2009) and Penner et al. (2011) in Figure 34 for comparison (as red circles and magenta triangle respectively). Our measurements are in good agreement with these previous values.

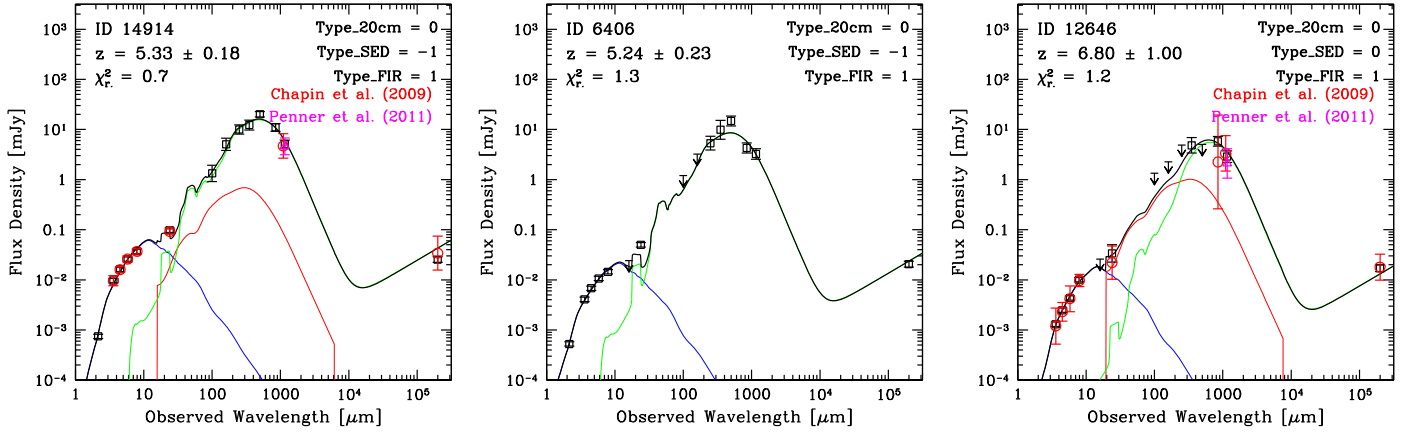
Note that the 3D-*HST* catalog (Skelton et al. 2014) reports a stellar mass of  $M_{*,3\text{D-HST}} = 10^{12.55} M_{\odot}$  for this object, at a photometric redshift  $z = 5.57$ . However, from our SED fitting, or simply by comparing IRAC fluxes with other, similar sources, such as GN20, we find a much lower stellar mass at a similar redshift,  $z = 5.33$ . (see footnote 36) This value is not far from our own SED fitting results, and it seems more reasonable than the 3D-*HST* value. Also note that this source is not in the *goodArea*; thus it was not included in our cosmic SFR density analyses.

ID 6406 is detected in the long-wavelength FIR and (sub)mm bands. Image cutouts are presented in Figure 36. It is almost

<sup>35</sup> The photometric redshift range in brackets indicate the  $1\sigma$  (68%) lower and upper confidence limits.

<sup>36</sup> GN20 has a peak stellar (over IRAC wavelengths) flux density of  $\sim 35 \mu\text{Jy}$ , redshift  $z = 4.055$ , and stellar mass  $M_{*} = 10^{11.28} M_{\odot}$  (see Figure 9). Scaling these values to  $z = 5.33$  with  $d_L^2 (1+z)^{-1}$  and to an IRAC peak flux of  $\sim 60 \mu\text{Jy}$  (see Figure 34), we would infer  $M_{*} \approx 10^{11.68} M_{\odot}$ .





**Figure 34.** The best-fit SEDs from the  $\langle U \rangle$ -dependent SED fitting for the three  $z \geq 5$  MS type FIR+mm galaxies: ID 14914, ID 6406 and ID 12646. The blue (near-IR wavelengths), red (if present at mid-IR wavelengths) and green (FIR wavelengths) curves represent the stellar, mid-infrared AGN (if present) and dust components. Parameters in the upper right corner are described in Section 3.1.

invisible in ACS images, but is detected in the near-IR WFC3, and becomes clear in the IRAC images, as well as the (sub)mm and radio data. Its 3D-*HST* and FIR+mm properties are listed below:

$$\begin{aligned}
 \text{ID}_{\text{IRAC}} &= 6406 \\
 \text{ID}_{3\text{D-HST}} &= 26337 \\
 z_{\text{FIR+mm}} &= 5.24 \pm 0.23 \\
 z_{3\text{D-HST}} &= 5.47 \text{ (3.725–5.82)} \\
 M_{*,3\text{D-HST}} &= 10^{11.44} M_{\odot} \\
 \text{SFR}_{3\text{D-HST}} &= 0.1 M_{\odot} \text{ yr}^{-1} \\
 \text{SFR}_{\text{FIR+mm}} &= 918 \pm 109 M_{\odot} \text{ yr}^{-1} \\
 \langle U \rangle &\approx 50.
 \end{aligned} \tag{8}$$

This source has no millimeter counterpart in Perera et al. (2008), Greve et al. (2008), and Penner et al. (2011). The closest mm source is AzGN31 in the Perera et al. (2008) catalog, at a distance of  $4''.9$ , and ID 33 in the Penner et al. (2011) catalog, at a distance of  $8''.8$ . Recent NOEMA mm follow-up observations detect a source that accurately corresponds with the IRAC and VLA 20 cm position (D. Liu et al. 2018, in preparation).

Multi-wavelength cutouts of ID 12646 are presented in Figure 37. There is no obvious detection in ACS or WFC3 images at the exact IRAC position. The closest 3D-*HST* source is 3D-*HST* ID 12669 at a photometric redshift  $z = 4.65$  (4.248–5.06), but this is located  $1''.52$  away from the IRAC position. Given the radio localization, this is unlikely to be the counterpart of ID 12646. ID 12646 is also not matched to any source in the catalog of Pannella et al. (2015); hence we do not have a stellar mass for it. The emission starts to be seen in the WIRCAM  $K_s$  image, but has an irregular morphology and low S/N ratio. It is fairly bright in IRAC images and bright at (sub) mm wavelengths, but with some offset from the emission peak. It matches AzGN10 in Perera et al. (2008), Chapin et al. (2009), and Penner et al. (2011). Its physical properties are listed below:<sup>37</sup>

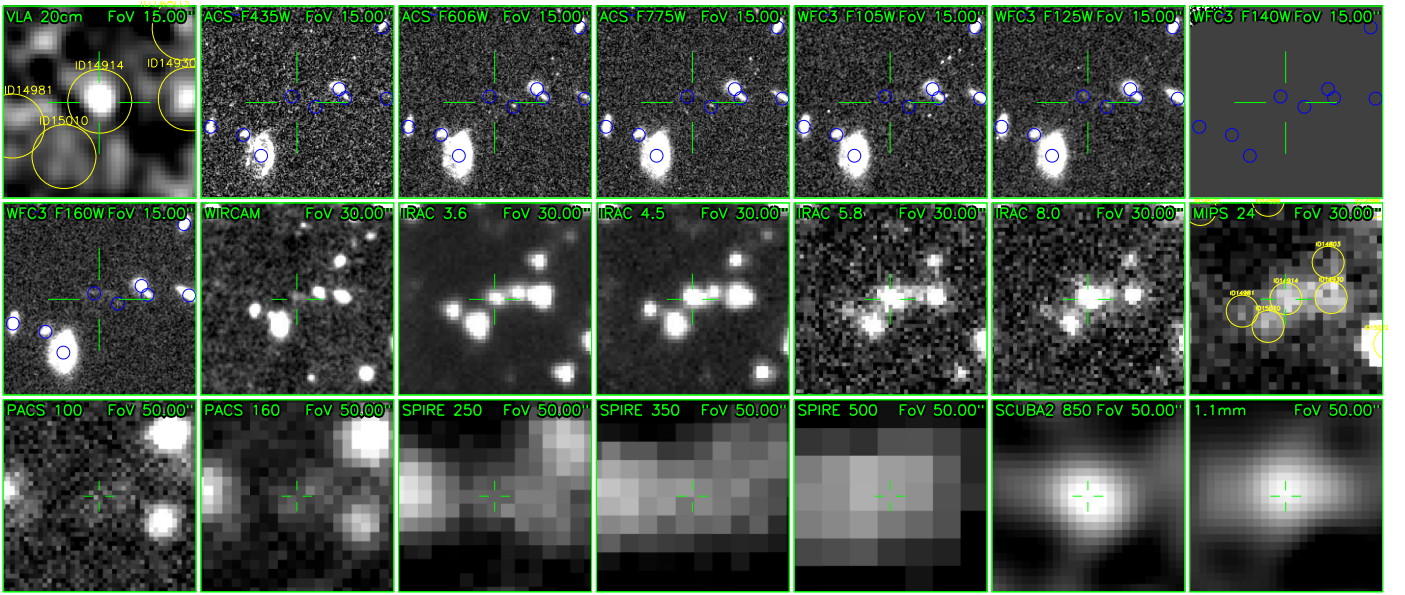
$$\begin{aligned}
 \text{ID}_{\text{IRAC}} &= 12646 \\
 z_{\text{FIR+mm}} &= 6.8 \pm 1.0 \\
 M_{*} &\approx 10^{11.44} M_{\odot} \\
 \text{SFR}_{\text{FIR+mm}} &= 814 \pm 249 M_{\odot} \text{ yr}^{-1} \\
 \langle U \rangle &\approx 31.
 \end{aligned} \tag{9}$$

The photometry of Chapin et al. (2009) and Penner et al. (2011) is also shown in the ID 12646 panel of Figure 34. The measurements generally agree with each other. To estimate the photometric redshift, Chapin et al. (2009) adopted a simple method which infers redshift as a function of IRAC and  $24 \mu\text{m}$  flux according to Pope et al. (2006), and derived a photometric redshift  $z \sim 3$ . Note that using their equation, GN20 (ID 564 in this work) with a spectroscopic redshift of 4.055 would have a photometric redshift  $z \sim 2.6$ .

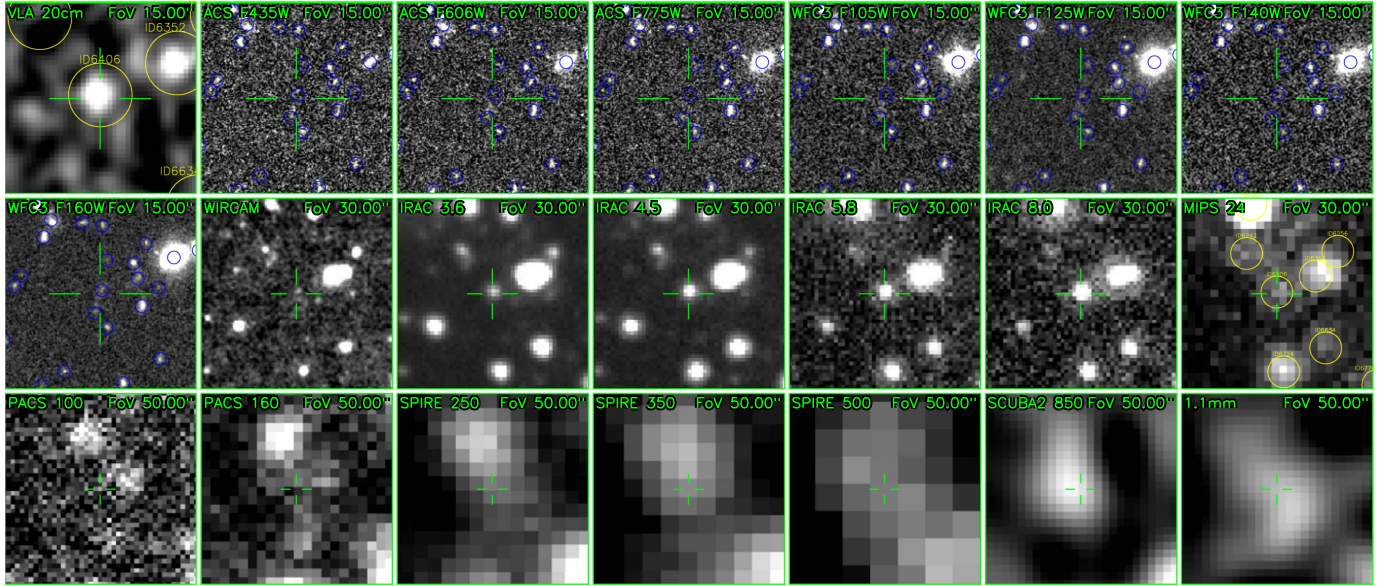
While these sources are interesting candidates for the most distant IR-detected galaxies in GOODS-N, we emphasize that we still only have photometric redshift estimates. In particular, the optical/near-IR SEDs are very red, faint at optical wavelengths, and rising at  $>3 \mu\text{m}$ . It is hard to obtain precise redshifts in these cases. However, note that the redshifts estimated from the FIR+mm SEDs agree well with the optical/near-IR photometric redshifts in all cases, while using templates with  $\langle U \rangle \sim 30\text{--}40$ , lower than what is expected for MS galaxies at  $z > 4$  (B  thermin et al. 2015). Because of the dust temperature versus redshift degeneracies, using warmer templates would further increase the estimated FIR+mm redshifts. We cannot exclude that these sources are actually SB galaxies, and their inferred  $\langle U \rangle$  values would be appropriate in that case. However, even in that case, substantially lower redshifts would require very cold dust SEDs, which seems unlikely.

Spectroscopic observations will be essential to measure redshifts for these objects, and to understand the reasons behind the different SED fitting results. Due to the dusty nature of these FIR+mm galaxies, optical/near-IR spectroscopic observations will likely not be as efficient as for LBGs. Mid-IR spectroscopic surveys with *JWST*, and (sub)mm spectral line scans with ground-based interferometric arrays, will be the most efficient ways to measure redshifts.

<sup>37</sup> Similarly, by scaling GN20 to a peak stellar flux of  $\sim 23 \mu\text{Jy}$  over IRAC wavelengths and a photometric redshift of 6.8, we would infer  $M_{*} \approx 10^{11.44} M_{\odot}$ .



**Figure 35.** Multi-wavelength cutouts of ID 14914. The instrument and bandpass of each cutout image is labeled in each panel. Yellow circles in the first panel and the MIPS 24 panel are our 24+radio sources and have a diameter of  $5''$ . Blue circles in the *HST* panels are the 3D-*HST* sources and have a diameter of  $0''.5$ . All panels are centered on the IRAC position (as indicated by the green cross) of the source under examination (ID 14914). The field of view is also indicated in each panel.



**Figure 36.** The multi-wavelength cutouts of ID 6406. See the caption for Figure 35.

## 9. Summary

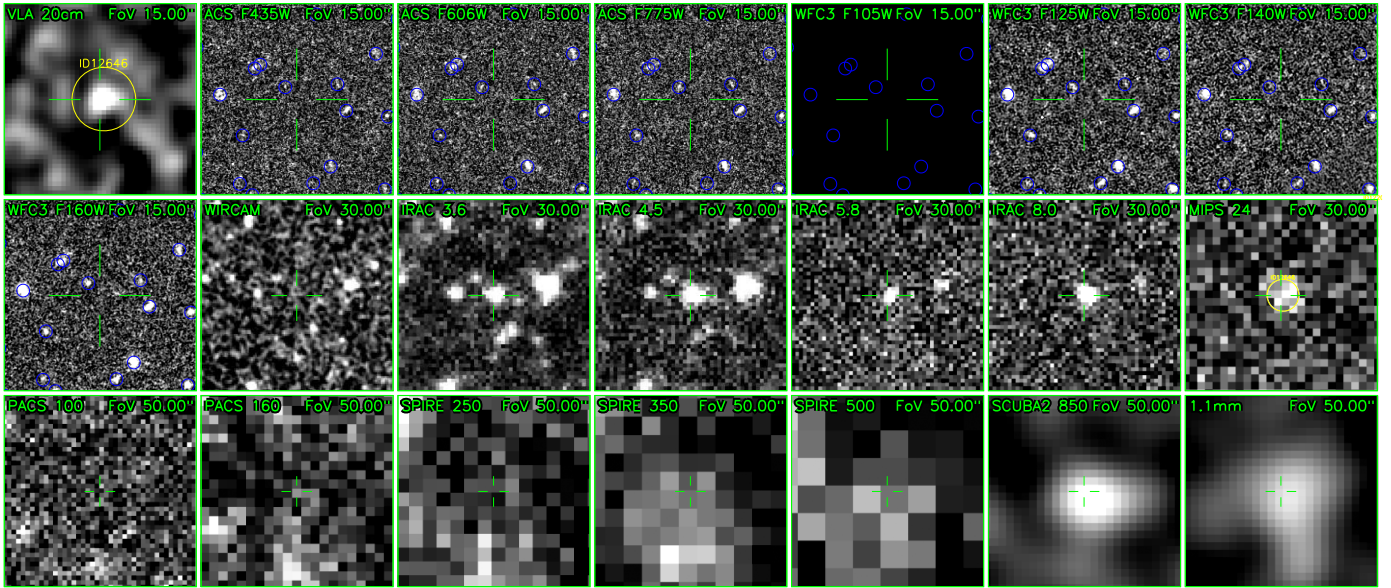
In this work, we have presented detailed super-deblended photometry for the exceptionally deep far-infrared to (sub-) millimeter imaging data that are available in the GOODS-North field. To overcome the heavy blending problems introduced by the large PSFs of these data, especially at *Herschel* SPIRE and (sub)mm wavelengths, we have developed a new method in which we choose the prior sources to use in fitting for the photometric measurements, and the ones to be “frozen” and subtracted at each FIR/mm band. In this method, we run SED fitting to predict the flux density of each source for each band, and then determine the

critical flux value for choosing an actual prior source list at each band by considering both the number density and the expected flux detection limit.

The number of fitted sources for each band can therefore be kept to reasonable values of  $\lesssim 1$  per PSF beam area. For the sources that are not fit, which are hopelessly faint at that wavelength, we eliminate their flux contribution by modeling their images and subtracting them from the observed data. In this way, the problem of flux boosting of fitted sources can also be largely reduced.

We generated Monte Carlo simulations to verify our photometric measurements, and to generate statistical correction recipes for obtaining reasonable measurement uncertainties





**Figure 37.** Multi-wavelength cutouts of ID 12646. See the caption for Figure 35.

with a nearly Gaussian behavior. The corrections are linked to three measurable parameters in three-step recipes. Thus, for real sources, we can correct flux biases and uncertainties for each source based on its measured parameters. The final uncertainties follow well-behaved statistics, which are important for obtaining more reasonable physical properties from SED fitting.

Finally, we identify a list of 1109 IRAC catalog sources with combined  $S/N \geq 5$  over the FIR and (sub)mm bands, including 70 detections at  $z \geq 3$ . Comparing with the empirical  $SFR-M_*$  MS relation at these redshifts, the majority of these dusty galaxies appear to be classified as “normal” MS galaxies, although we caution that the MS level is not yet very well known at these highest redshifts.

We compute the co-moving SFR densities that are directly observed for our detected FIR+mm sources using the  $1/V_{\max}$  method plus a boot-strapping analysis. We estimate the completeness in SFR for our FIR+mm sample from the SMF-converted SFR histograms. At  $z \leq 2$ , values estimated using different SMFs from the literature are similar and consistent, but they vary considerably at  $z \sim 2-4$  due to the existence of significant disagreements between faint-end SMF slopes in the literature. These are even worse at  $z \sim 4-6$ , where we effectively detect only the few most massive galaxies, including, notably, galaxies in the GN20 proto-cluster at  $z = 4.05$ .

We find that the completeness-corrected cosmic SFR densities agree well with the standard literature results from Madau & Dickinson (2014) and Gladders et al. (2013) at  $z \leq 2$ . At  $z \sim 2-4$ , the corrections vary strongly with the adopted SMFs, but we suggest that the total SFR density could be higher than the literature SFR densities by as much as a factor of 3. At  $z > 4$ , we cannot reliably estimate the incompleteness, so we report only the directly observed SFR densities in Figure 28, and that value minus the  $1\sigma$  uncertainty as lower limits to the dust-obscured cosmic SFR in Figure 30.

We find that our estimates for co-moving SFR densities are largely independent from those derived from UV-selected galaxies, and thus should be in principle summed with the

UV-derived values, at least for  $z \gtrsim 3$  and only for the directly measured quantities at the bright end, as our sample does not obviously overlap with  $B$ - and  $V$ -dropout LBGs.

We finally present the discovery of three galaxies expected to be at  $z > 5$ , including one at  $z \sim 6.8$ , that are the most distant candidates in our GOODS-N sample. Their stellar masses are seemingly very large, a peculiarity shared with much of the sample at  $z > 4$  that merits further investigations.

We are grateful to the referee for a detailed and helpful report that resulted in an improved presentation of this paper. We thank Iary Davidzon for helpful discussion on stellar mass functions, and Daniel Stern and the late Hyron Spinrad for unpublished spectroscopic redshifts that are used in this paper. D.L. and E.D. would like to thank the Chinese Academy of Sciences (CAS)—Centre national de la recherche scientifique (CNRS) Joint PhD Program. Y.G. acknowledges support from the National Natural Science Foundation of China (grants 11390373 and 11420101002) and the Chinese Academy of Sciences’ Key Research Program of Frontier Sciences. M.T.S. acknowledges support from a Royal Society Leverhulme Trust Senior Research Fellowship (LT150041). This work is based in part on observations made with Herschel, a European Space Agency Cornerstone Mission with significant participation by NASA. Support for work by M.E.D. and H.I. was provided by NASA through an award issued by JPL/Caltech. Their work was also supported by a NASA Keck PI Data Award, administered by the NASA Exoplanet Science Institute. Data presented herein were obtained at the W. M. Keck Observatory from telescope time allocated to the National Aeronautics and Space Administration through the agency’s scientific partnership with the California Institute of Technology and the University of California. The Observatory was made possible by the generous financial support of the W. M. Keck Foundation. The authors wish to recognize and acknowledge the very significant cultural role and reverence that the summit of



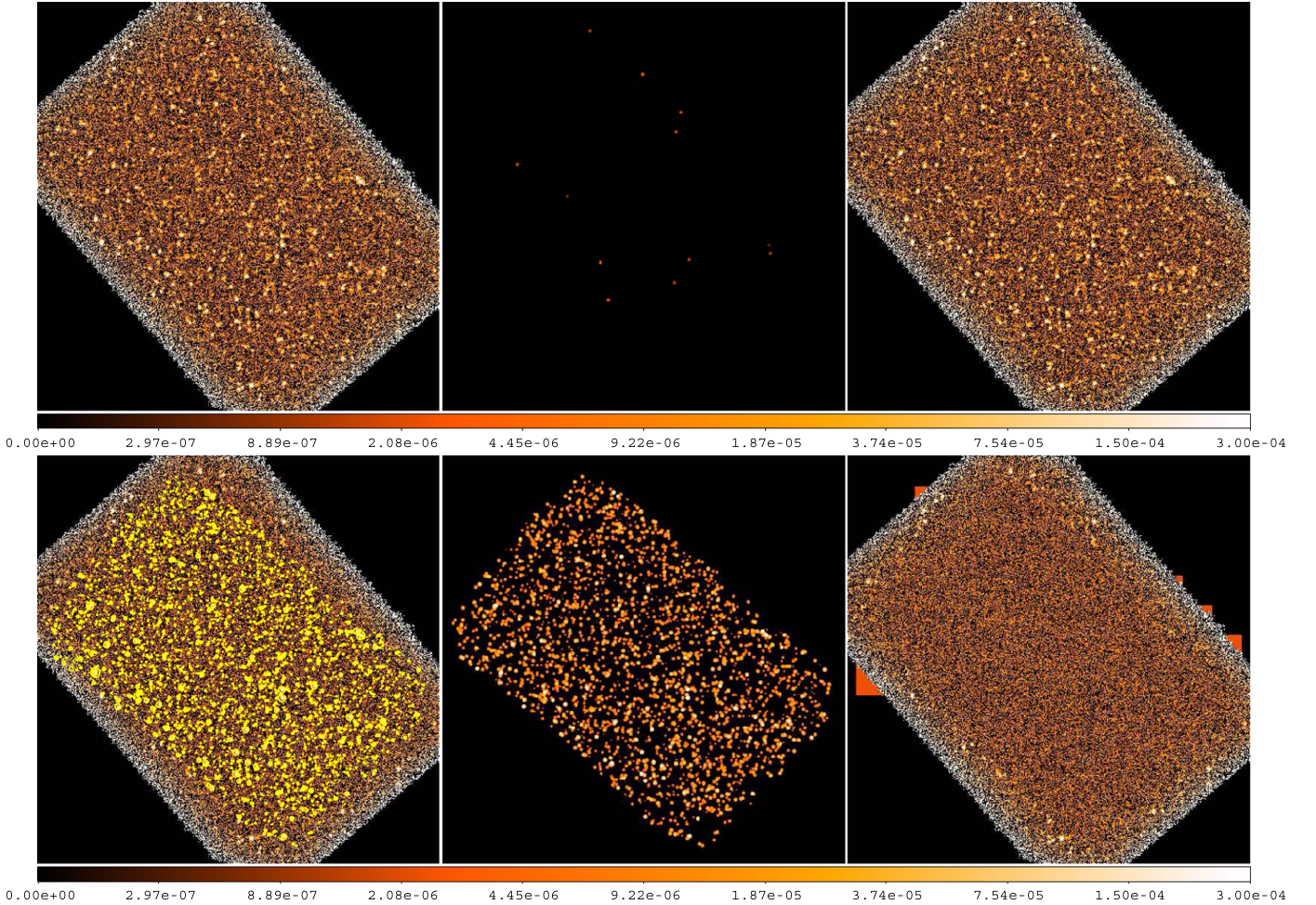
Mauna Kea has always had within the indigenous Hawaiian community. We are most fortunate to have the opportunity to conduct observations from this mountain.

### Appendix A Photometry Image Products

Here we present the photometry image products for all the bands analyzed in this paper. In Figure 38 (and all following figures in an analogous manner), the three images in the first row (top) demonstrate the process of subtraction of *hopelessly faint* sources (see description in Section 3.3). From left to right, we show the original image for this band, the faint-source-model image that is constructed based on SED-predicted flux values, and the faint-source-subtracted image that is made by subtracting the middle panel from the left panel. The three images in the second row (bottom) demonstrate the prior-extraction photometry for this band. From left to right, they show the

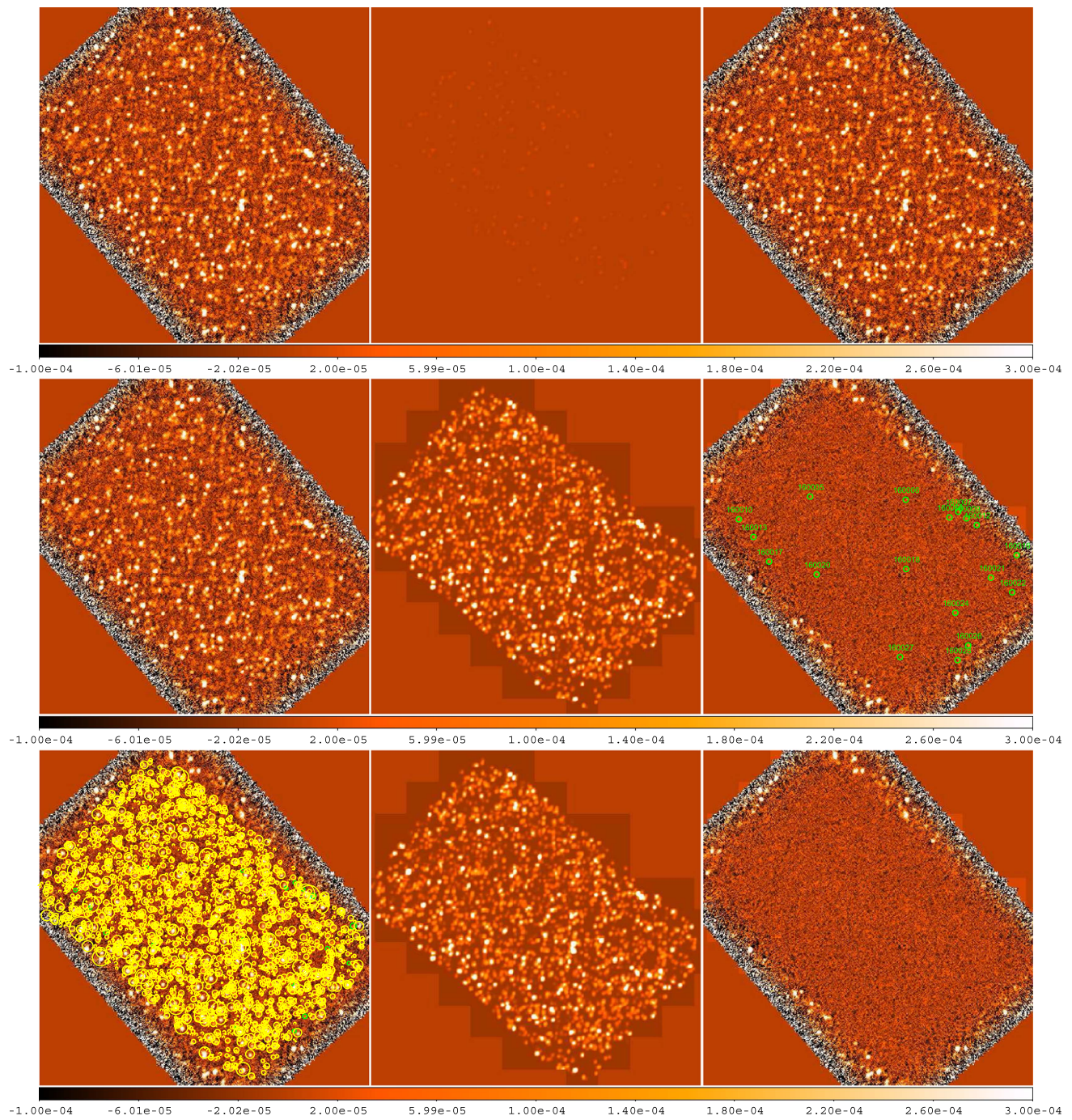
faint-source-subtracted image (identical to the top-right panel, but here marking sources retained to be fit with yellow circles), the best-fit-model image, and the residual image which is the left panel minus the middle panel. We run SExtractor on the residual image to detect additional sources for each band, shown with green circles (although, at this  $100\ \mu\text{m}$  band, we detect no additional source). The first two rows of Figures 39–44 are identical to the panels shown in Figure 38. The third row in these later figures, from left to right, shows the faint-source-subtracted image, the best-fit-model image (this time including both the prior sources and additional sources detected in the residual maps), and the final residual image.

All images have the same logarithmic scaling. The radii of the circle symbols are proportional to the square root of their measured source flux densities, but capped to a maximum when S/N reaches 10. Sources with  $S/N < 3$  are shown as dashed circles.



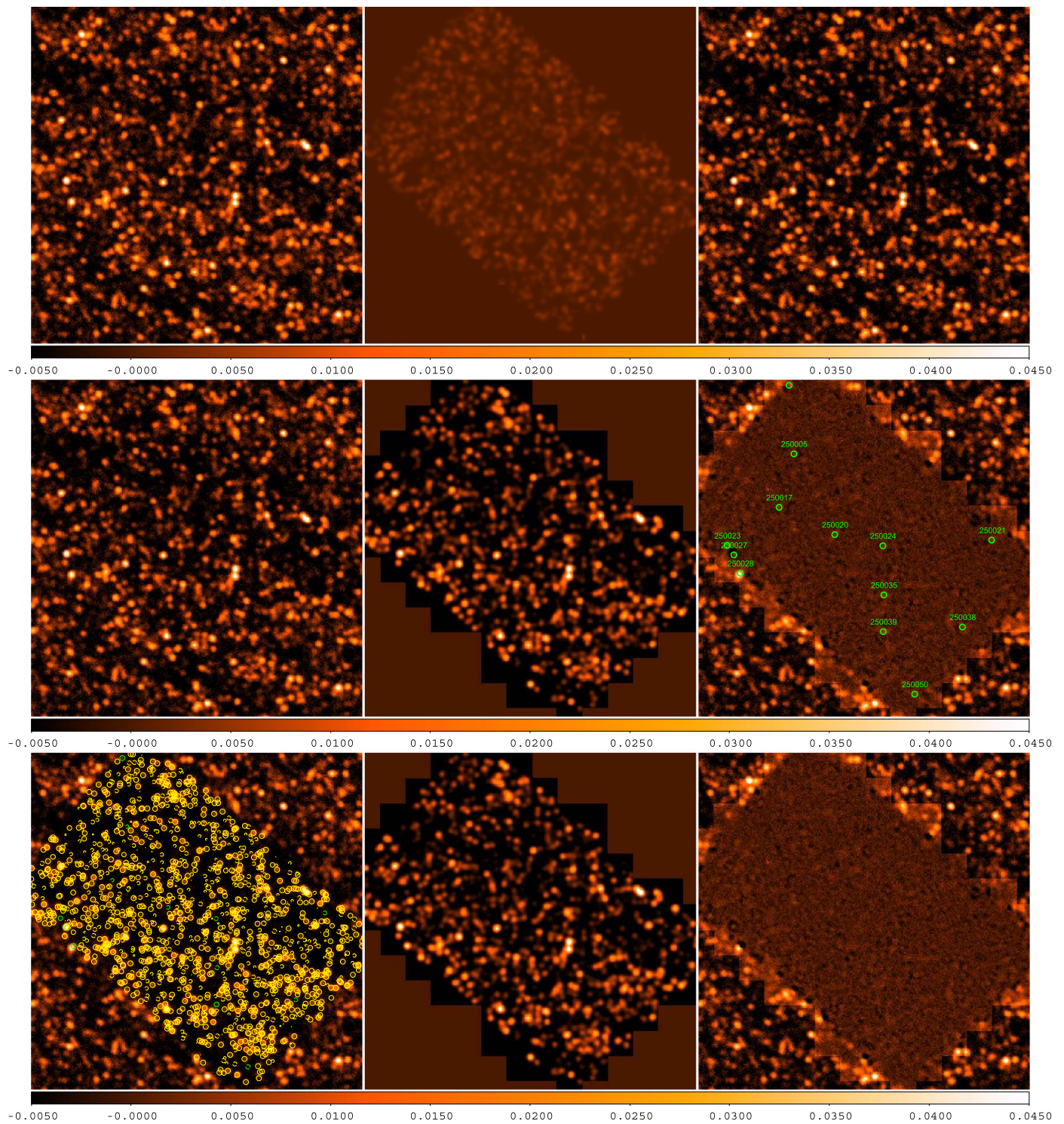
**Figure 38.** Photometry image products at  $100\ \mu\text{m}$ . See descriptions in the text.





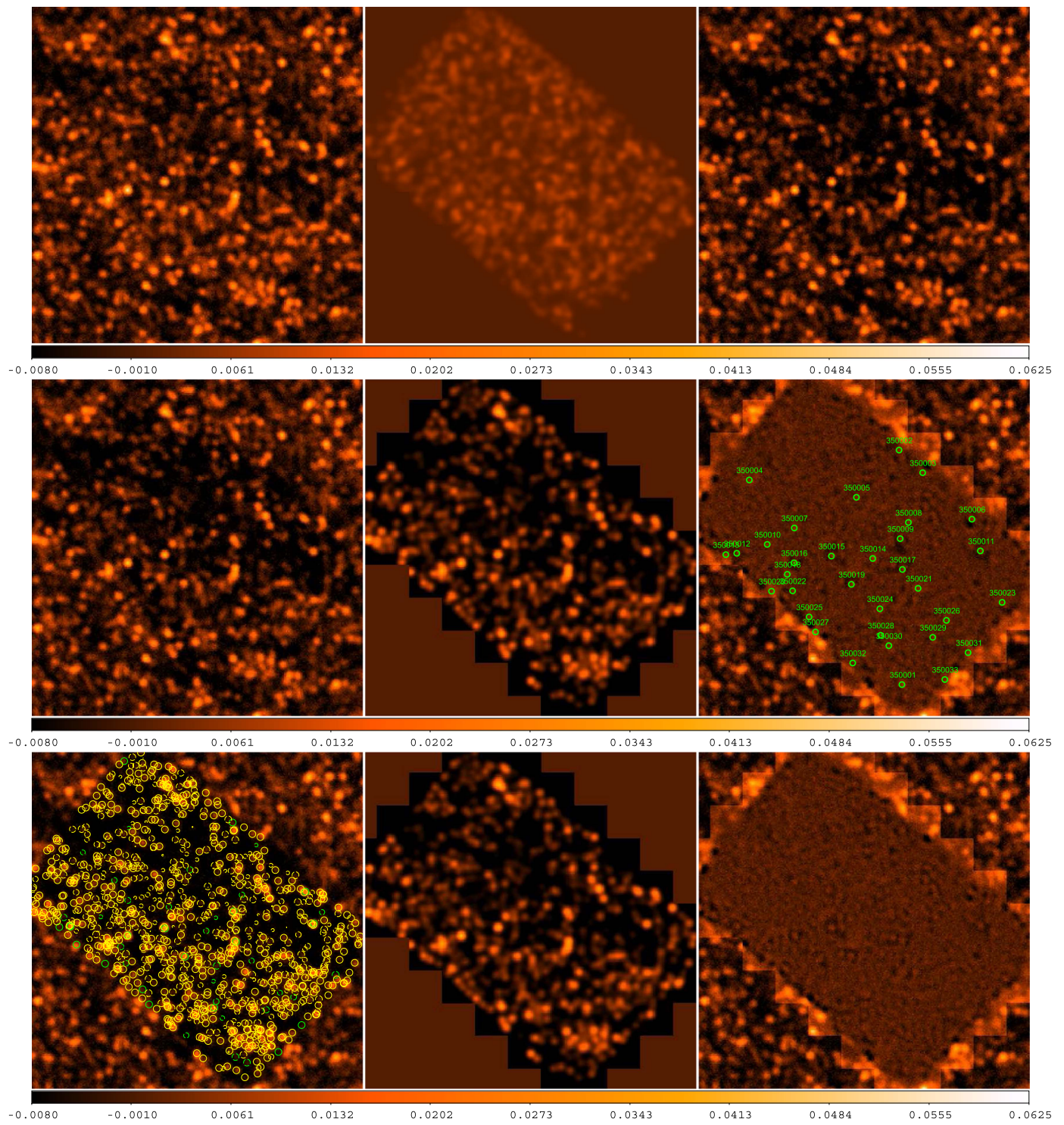
**Figure 39.** Photometry image products at  $160\ \mu\text{m}$ . See descriptions in the text.





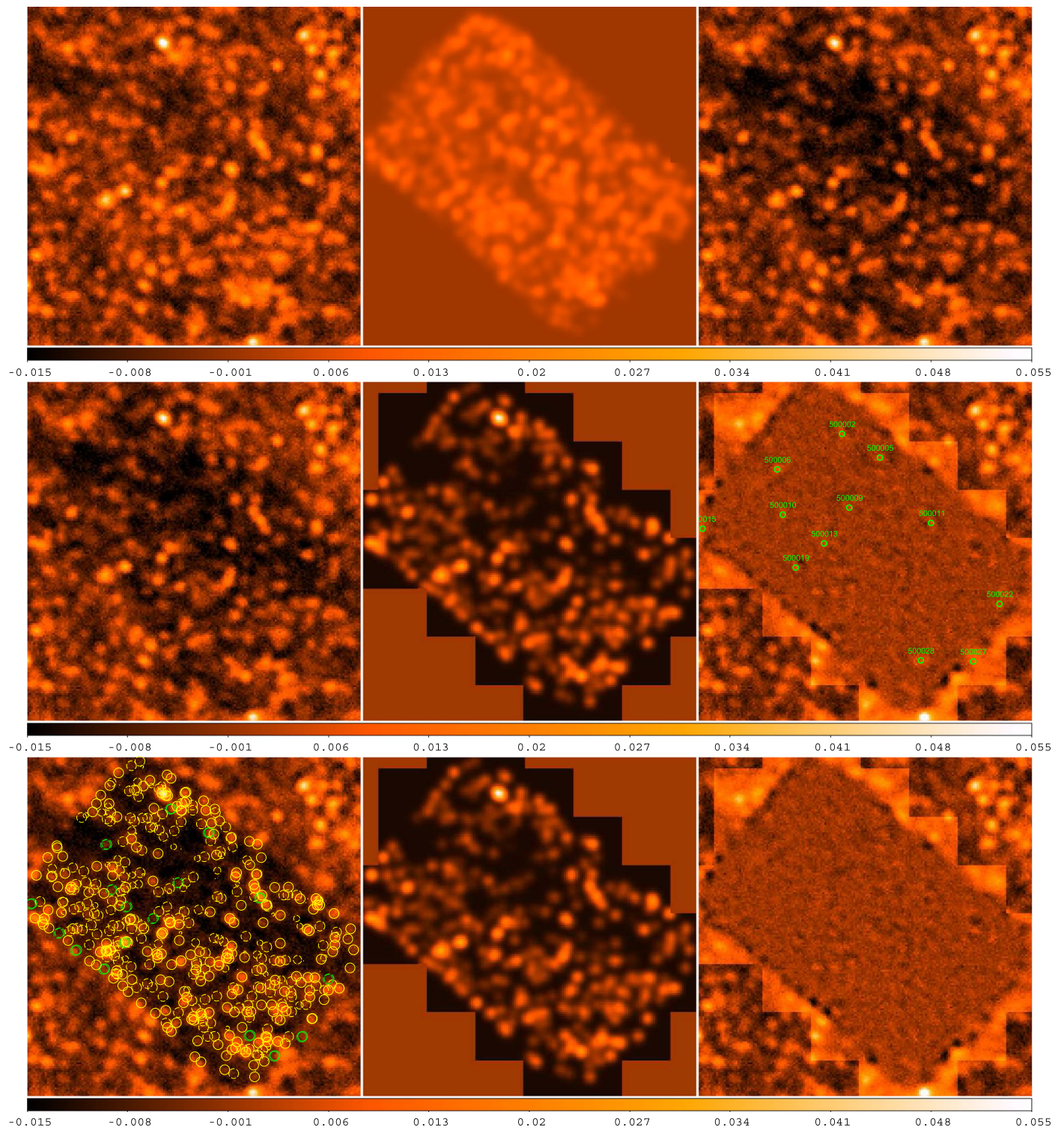
**Figure 40.** Photometry image products at  $250\ \mu\text{m}$ . See descriptions in the text.





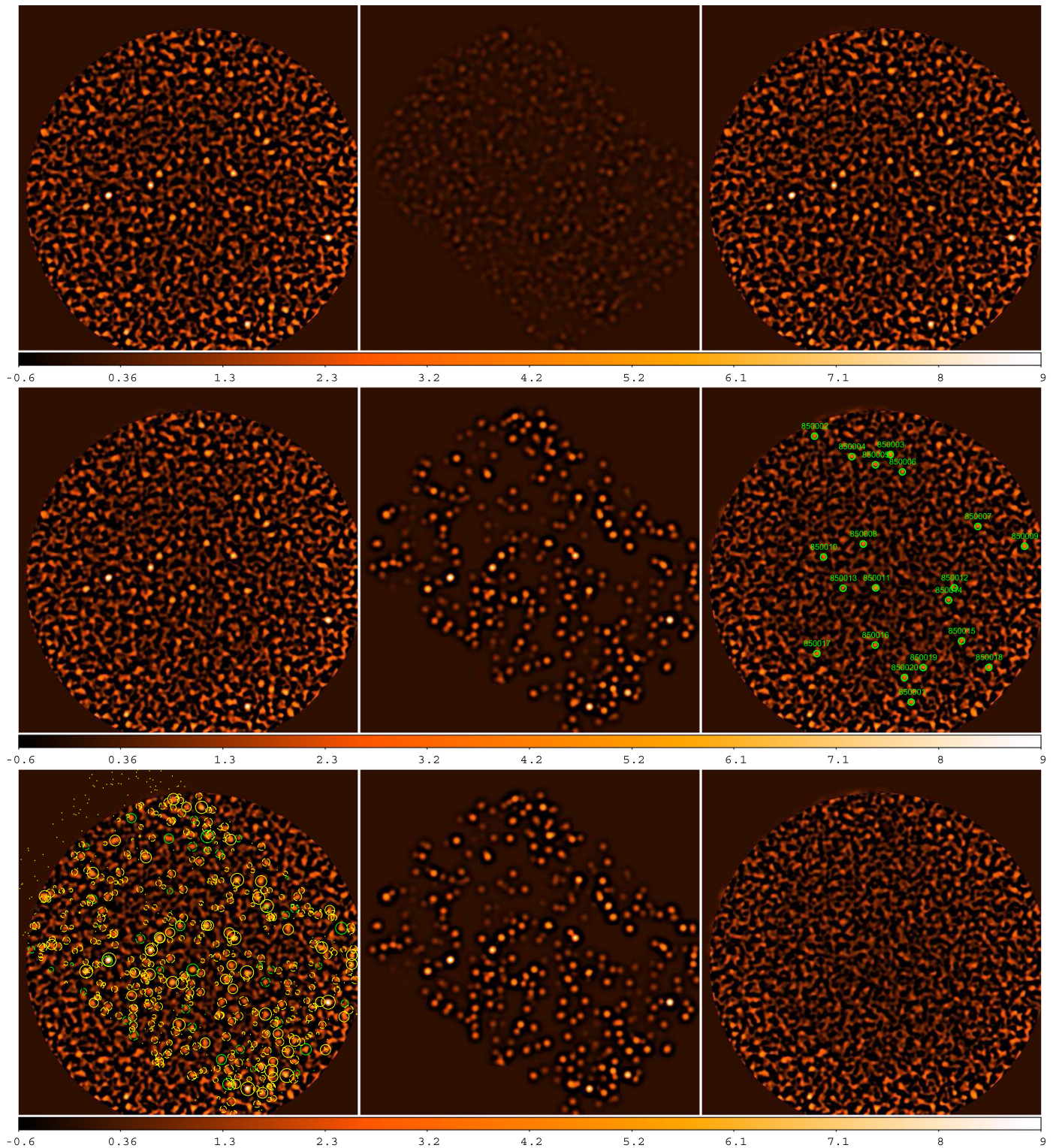
**Figure 41.** Photometry image products at  $350\ \mu\text{m}$ . See descriptions in the text.





**Figure 42.** Photometry image products at  $500\ \mu\text{m}$ . See descriptions in the text.





**Figure 43.** Photometry image products at 850  $\mu\text{m}$ . See descriptions in the text.

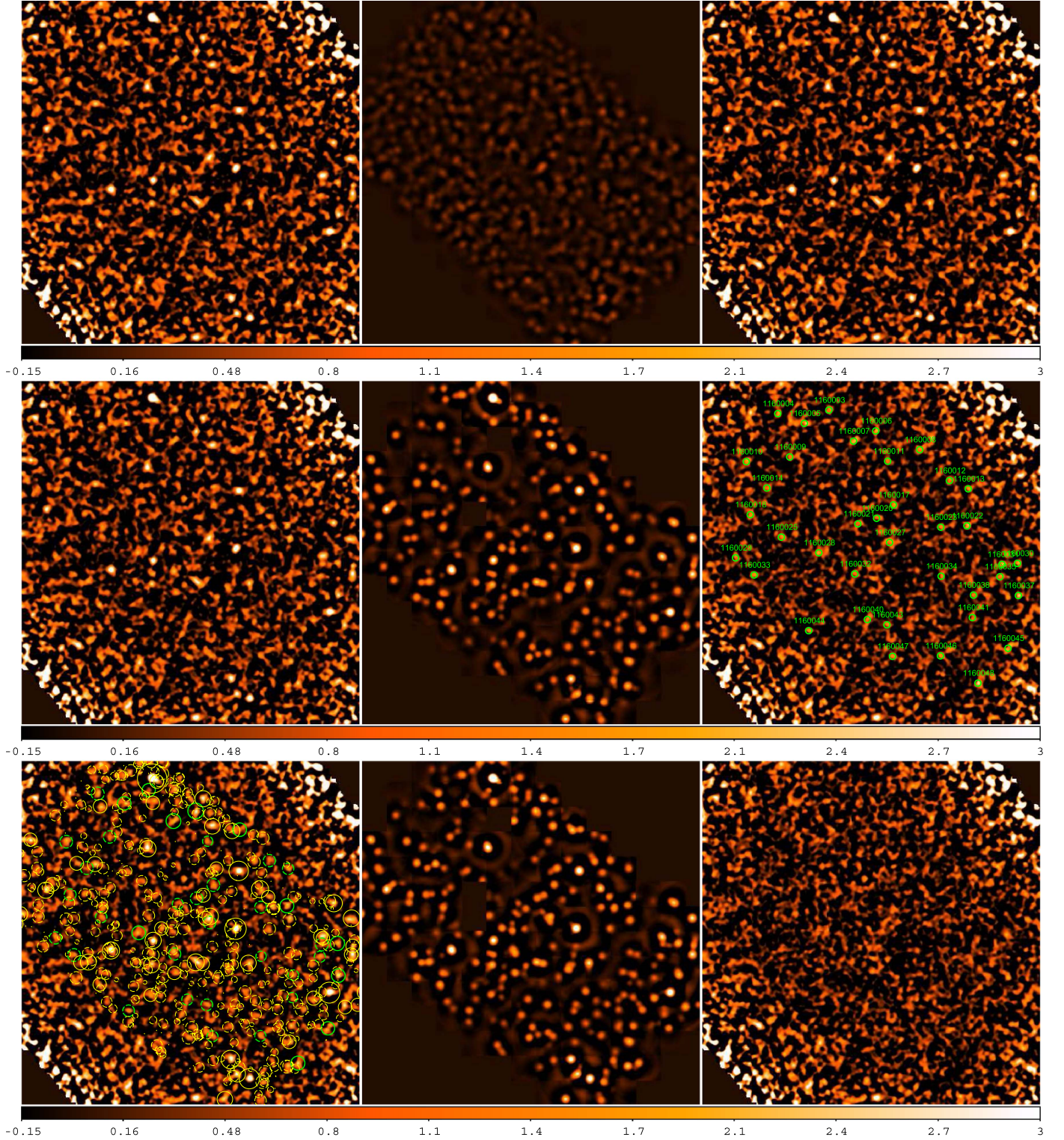
## Appendix B

### Simulation Correction Analyses

Here we present figures reporting the results of our simulation-based correction recipes for all the bands that we analyzed. In the main text, for simplicity, we have shown only

the case of 350  $\mu\text{m}$  (see Figure 11), and we do not duplicate that figure here. For example, in Figure 45, for the observations at 100  $\mu\text{m}$ , we bin the simulation data points by three measurable parameters: the *galfit* flux uncertainty normalized by the local rms noise at the source position ( $\sigma_{\text{galfit}}/\sigma_{\text{rms noise}}$ ) in





**Figure 44.** Photometry image products at 1.16 mm. See descriptions in the text.

the first column, the residual flux within one PSF beam area on the residual image, normalized also by the local rms noise ( $S_{\text{residual}}/\sigma_{\text{rms noise}}$ ) in the second column, and the *crowdedness* parameter (see Section 5.1) in the third column. The bins used

for the analysis are indicated by the dashed vertical lines in the first and second row images.

In the first row, we analyze the difference between the input and output fluxes of each simulated source ( $S_{\text{in}} - S_{\text{out}}$ ), which

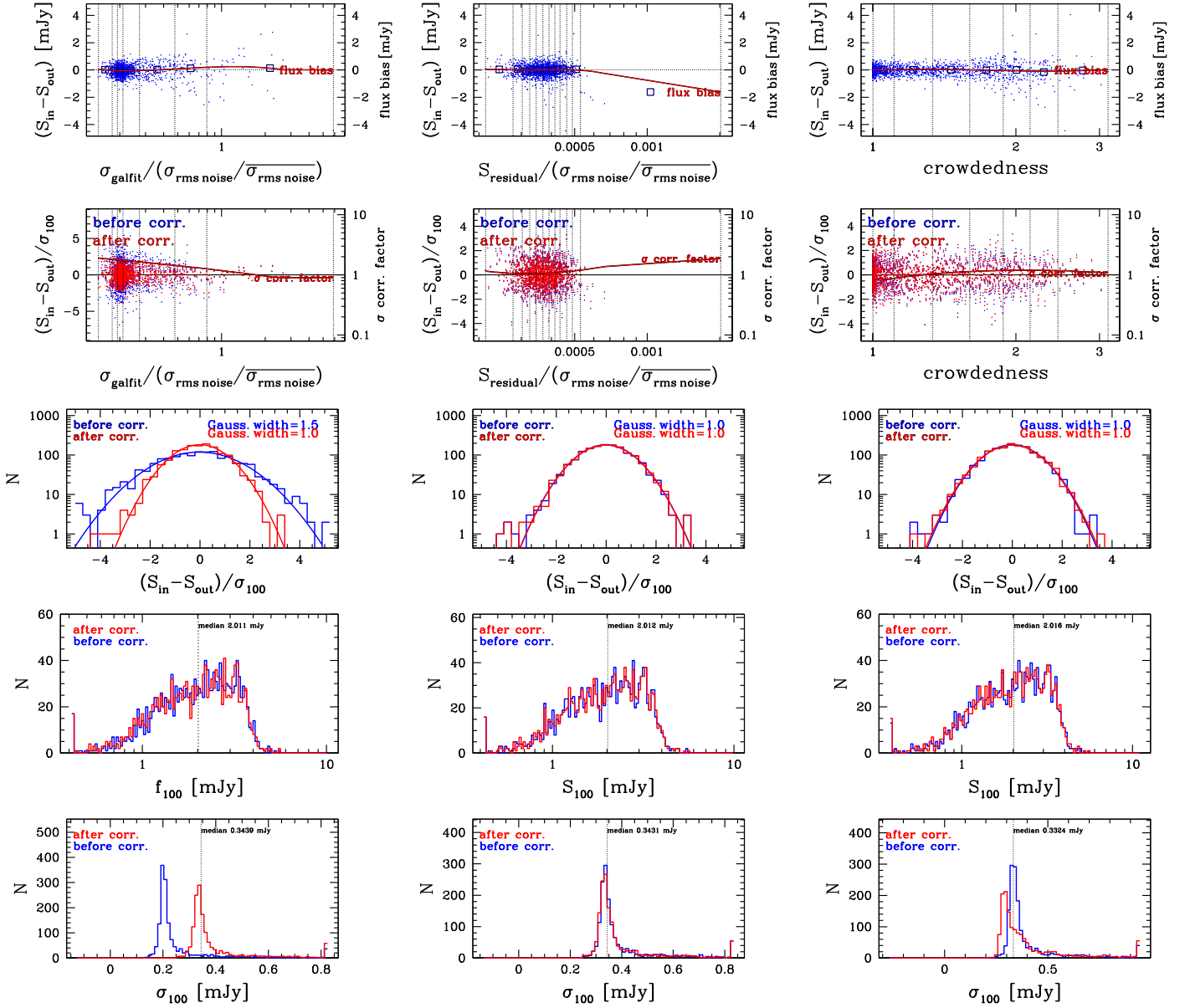


Figure 45. Simulation correction analyses at 100  $\mu\text{m}$ . See descriptions in the text.

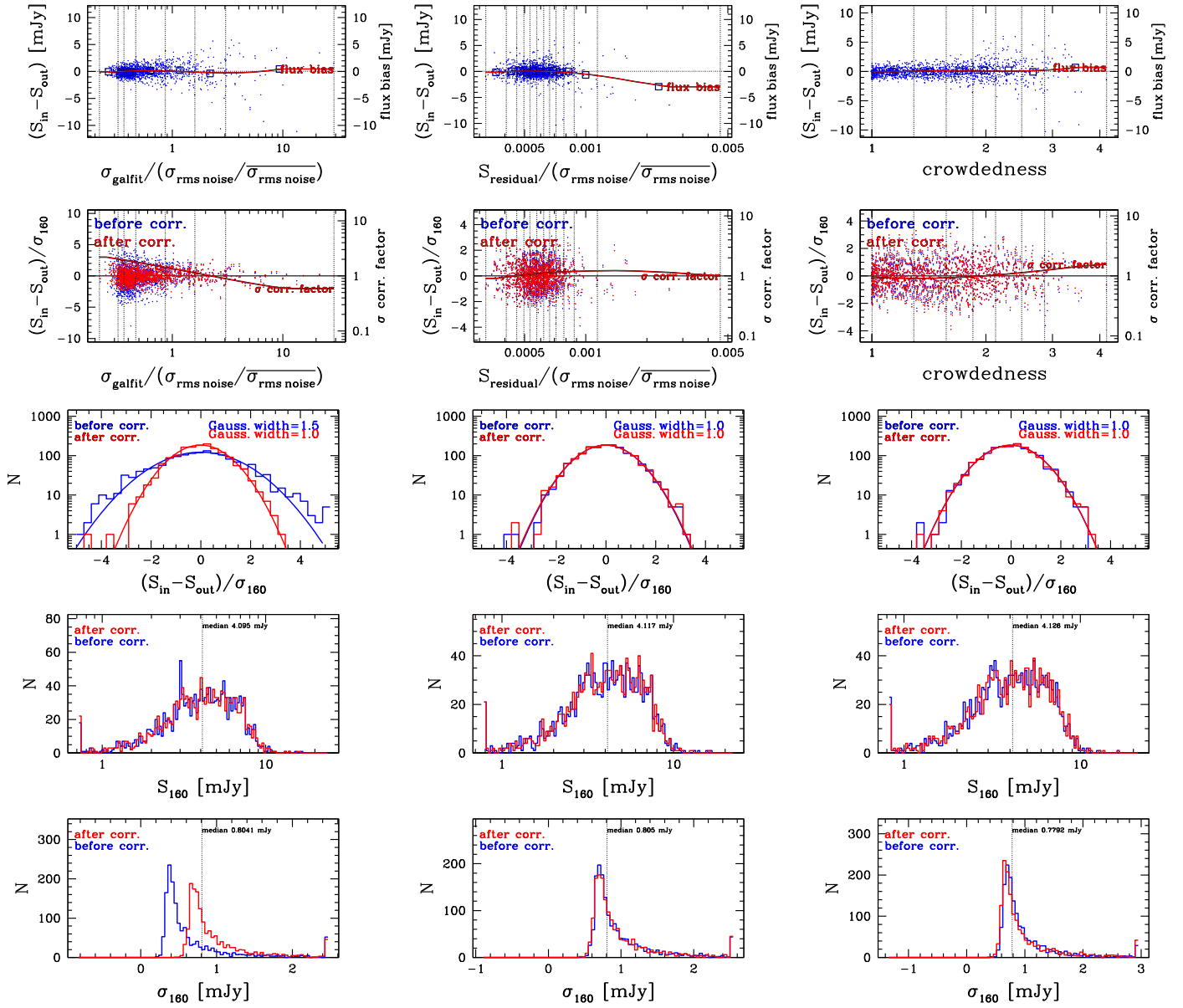
is used to correct the flux bias for our photometry. We fit a three-order polynomial function to the bin-averaged  $(S_{\text{in}} - S_{\text{out}})$  for each parameter, shown as the red curve.

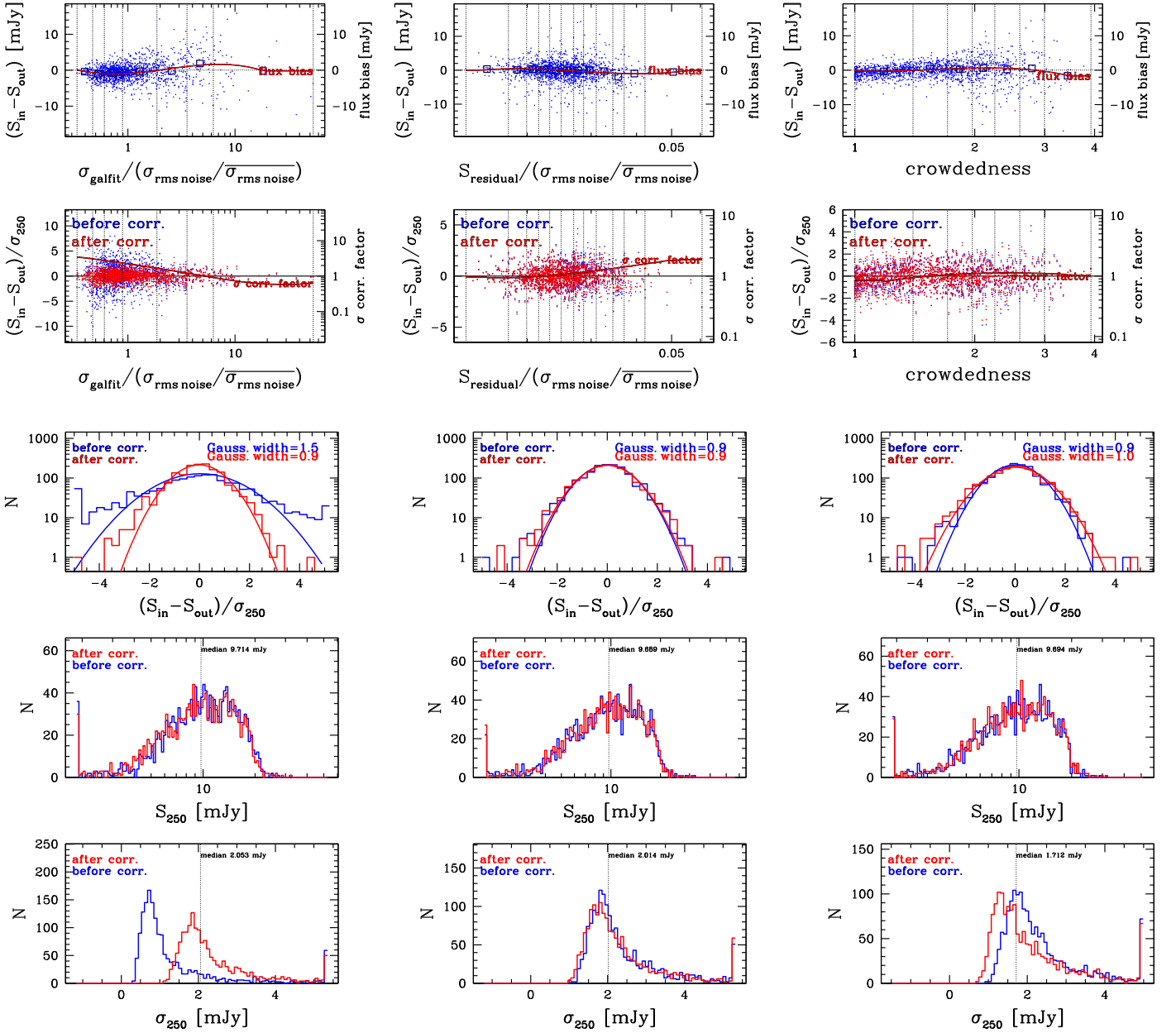
In the second row, we analyze the flux difference  $(S_{\text{in}} - S_{\text{out}})$  divided by the flux uncertainty ( $\sigma$ ) of each simulated source  $((S_{\text{in}} - S_{\text{out}})/\sigma)$ . The scatter in each bin is the correction factor to be applied to  $\sigma$ . Blue points show the uncorrected data points, while red points show the corrected values. After correction, the scatter of  $(S_{\text{in}} - S_{\text{out}})/\sigma$  in each bin is very close to 1.0, indicating that the corrected  $\sigma$  is statistically consistent with the scatter of  $S_{\text{in}} - S_{\text{out}}$ . We fit a three-order polynomial function to the bin-averaged flux uncertainty correction factor for each parameter, as shown by the red curve, whose value can be read on the right axis.

In the third row, we show the histograms of  $(S_{\text{in}} - S_{\text{out}})/\sigma$  before and after the three-parameter corrections. Its shape, after correction (i.e., the red histogram), get closer to a well-behaved Gaussian distribution (i.e., symmetric and with a Gaussian width of 1.0), and is thus an improvement over the uncorrected one (i.e., the blue histogram). See the text in Sections 5.3 for more detailed information.

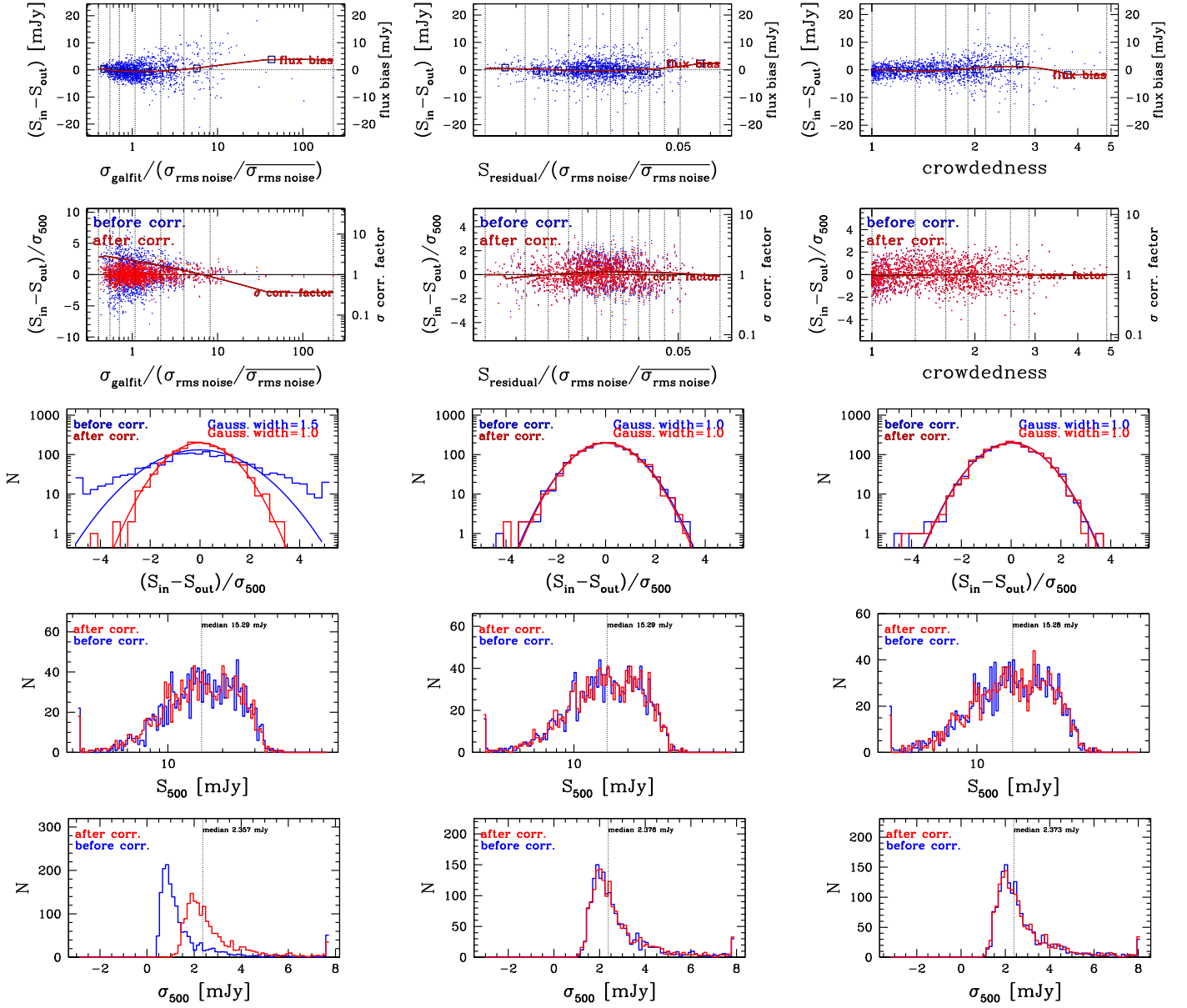
In the fourth and fifth rows, we show the histograms of flux densities, and of flux density uncertainties, respectively. In each panel, the blue histogram shows values before correction, and the red histogram shows the results after correction. From left to right, the panels are in the same three-parameter order. Generally the flux density changes are very small, while the corrections to the flux density uncertainties are larger with respect to the *galfit* initial values.

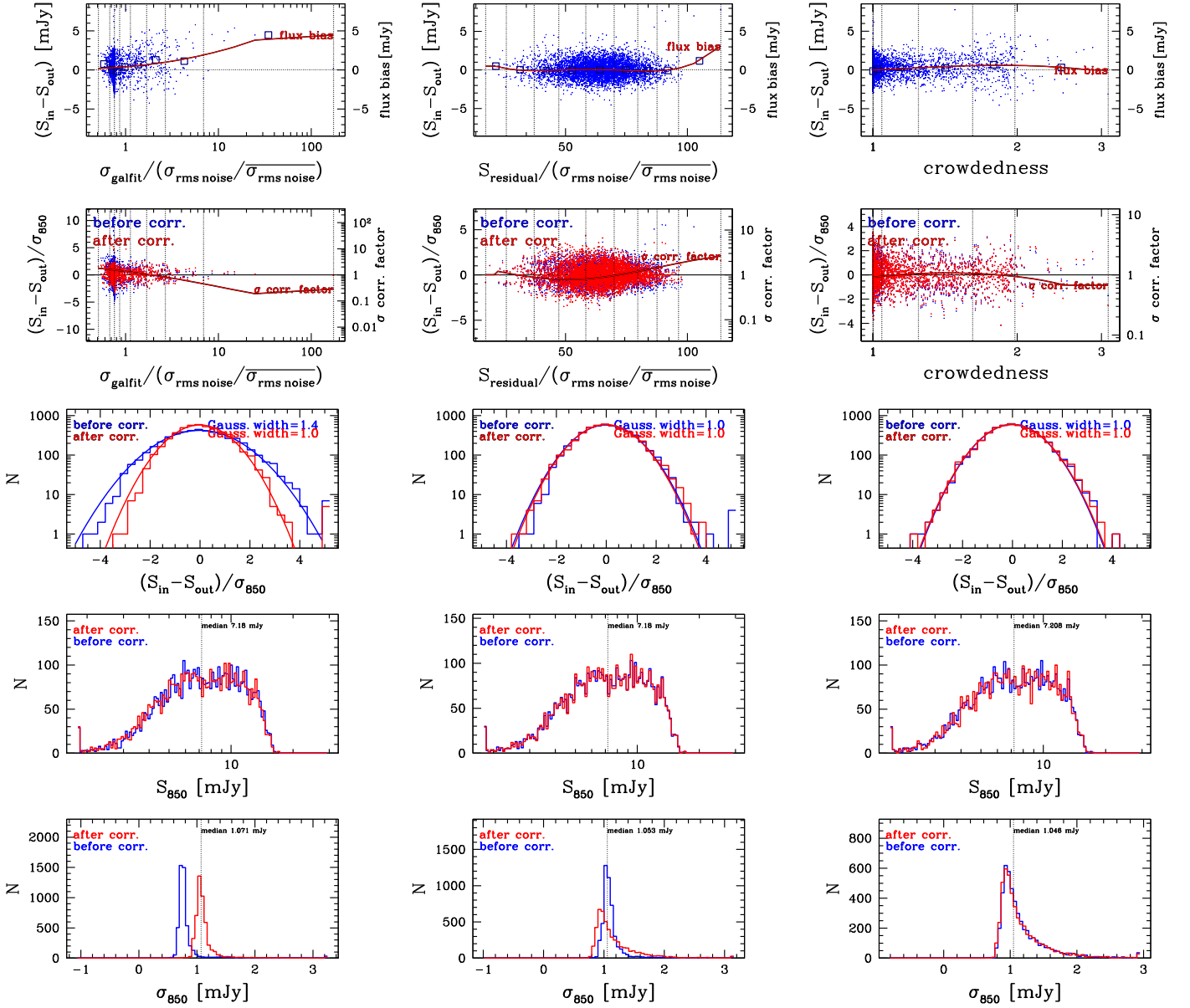


Figure 46. Simulation correction analyses at 160  $\mu\text{m}$ . See descriptions in the text.

Figure 47. Simulation correction analyses at 250  $\mu\text{m}$ . See descriptions in the text.



Figure 48. Simulation correction analyses at 500  $\mu\text{m}$ . See descriptions in the text.

Figure 49. Simulation correction analyses at 850  $\mu\text{m}$ . See descriptions in the text.

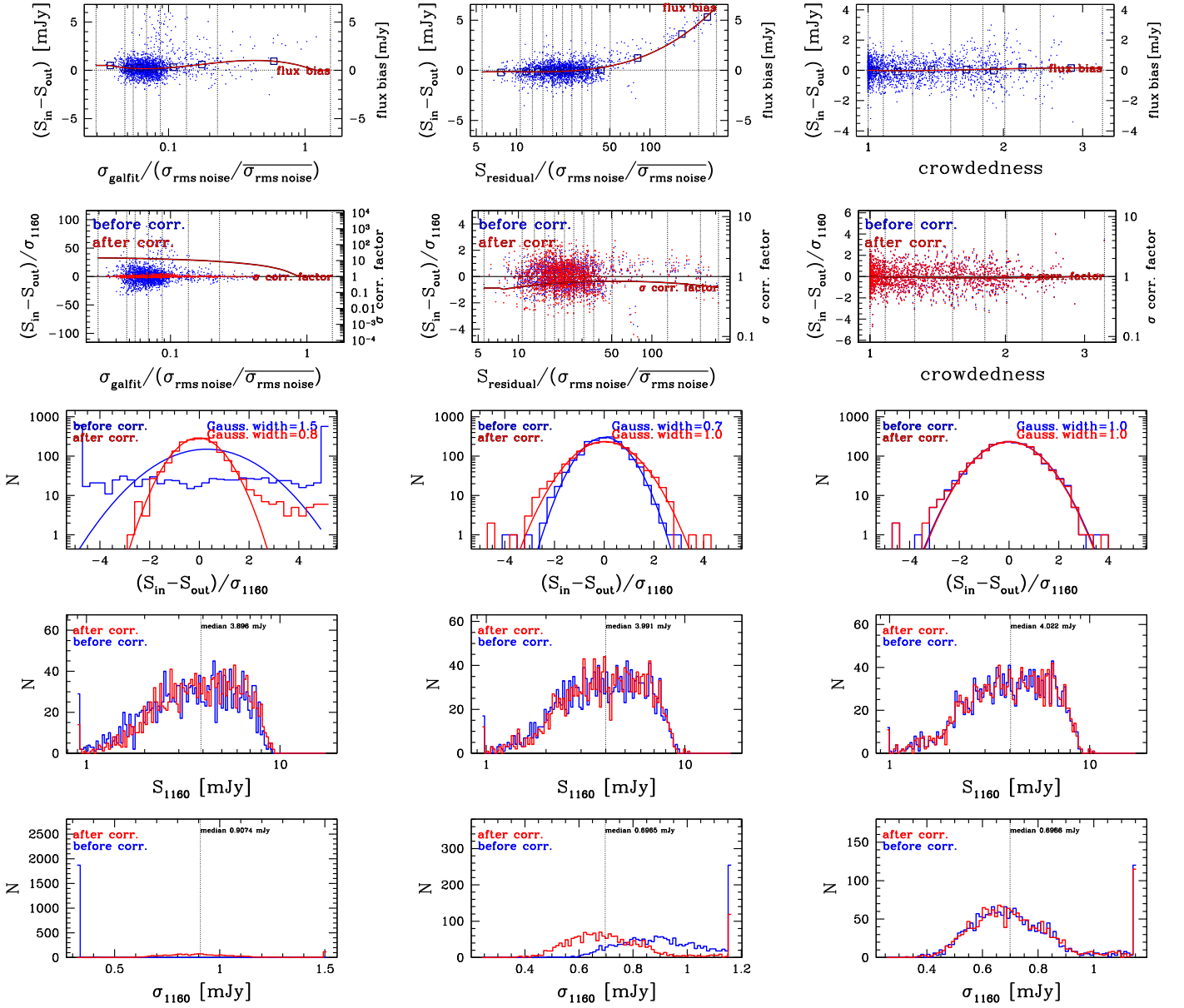


Figure 50. Simulation correction analyses at 1.16 mm. See descriptions in the text.



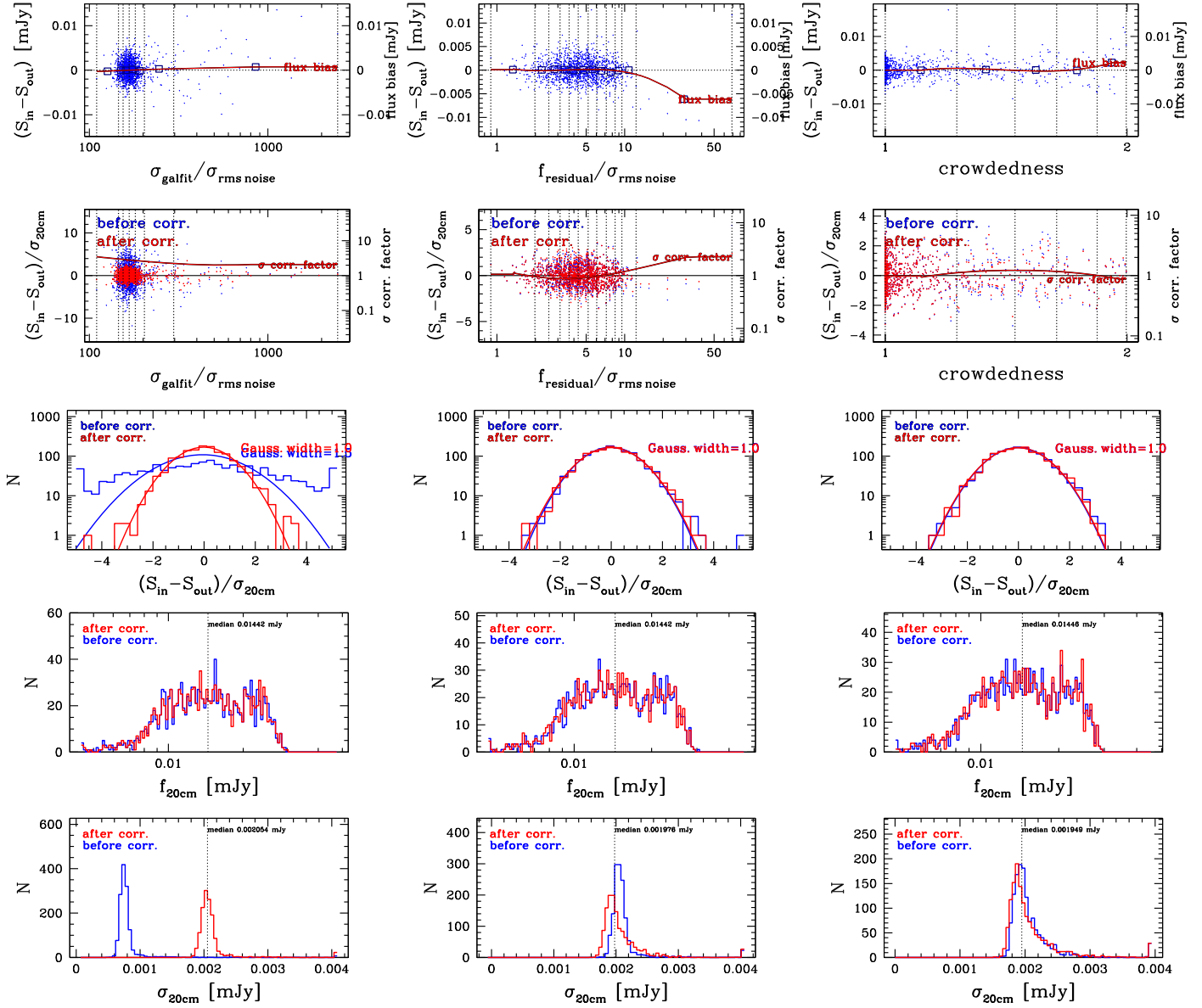
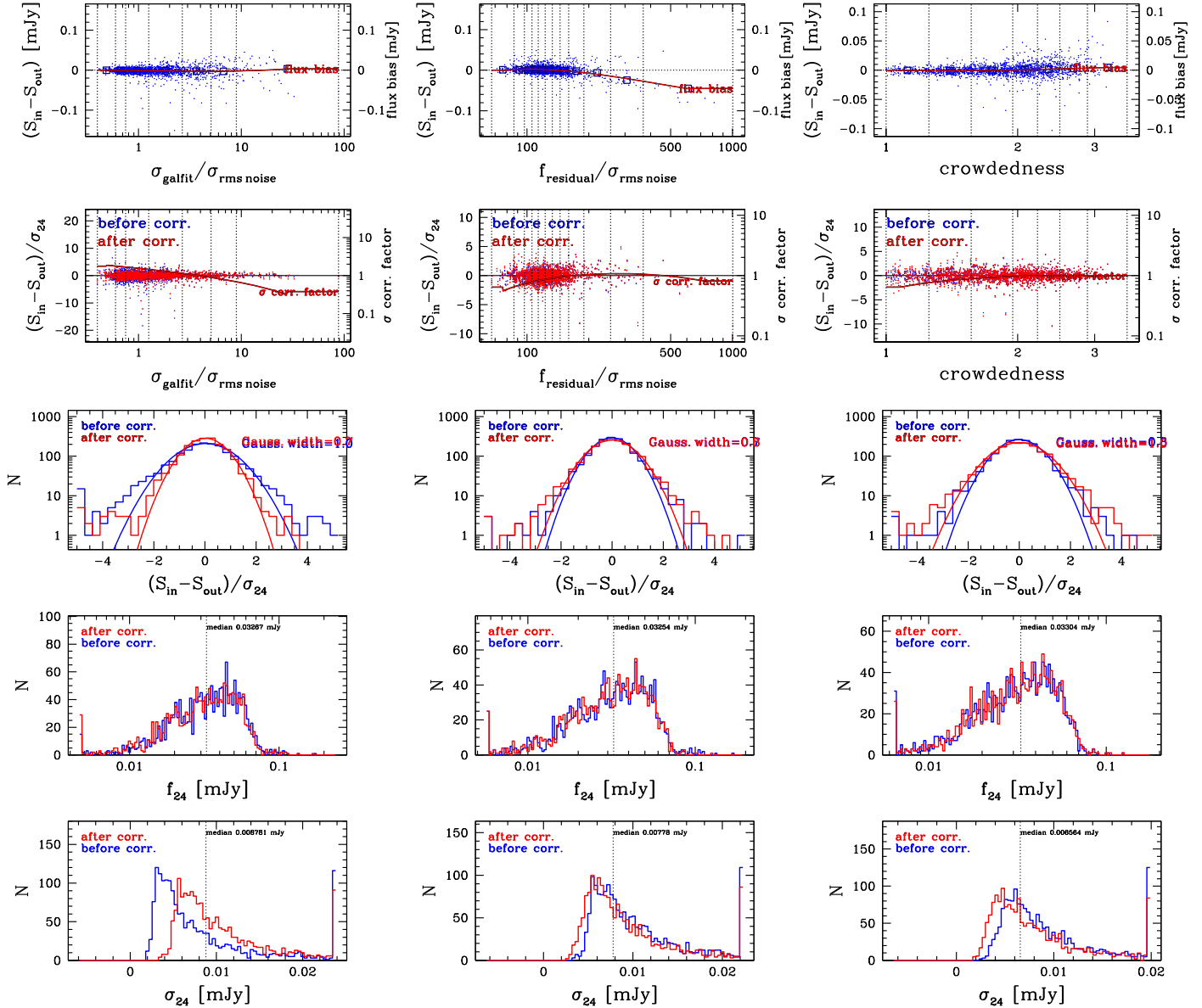


Figure 51. Simulation correction analyses at 20 cm. See descriptions in the text.

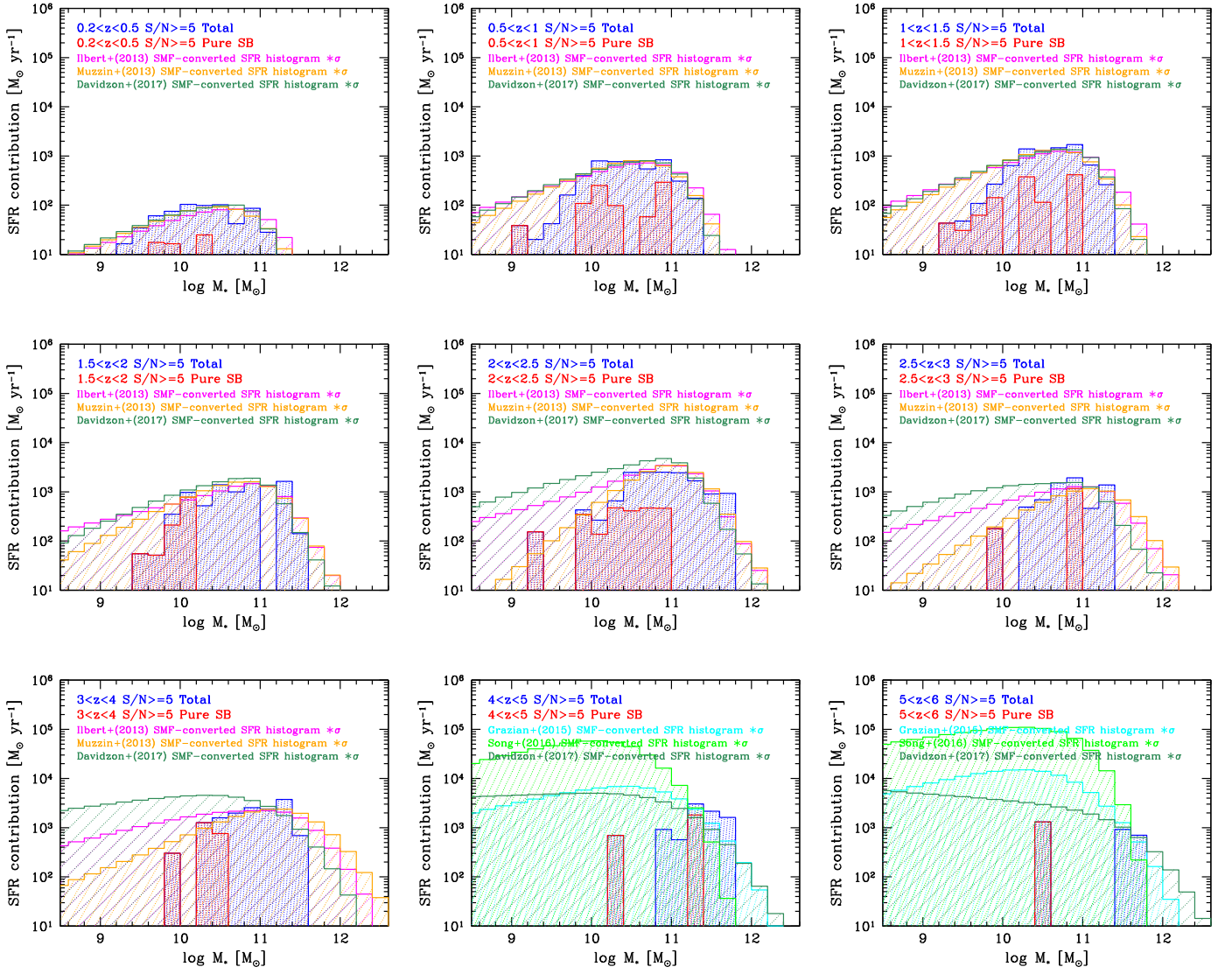
Figure 52. Simulation correction analyses at 24  $\mu\text{m}$ . See descriptions in the text.

### Appendix C

#### The Cosmic SFR Density Completeness Estimation Using SMF-converted SFR Histograms

In this work we use a new method to estimate the completeness of our FIR+mm sample by comparing our measurements with literature SMFs convolved with MS-like




SFR distributions. We compute the SMF-converted SFR histograms as described in Section 8.4, and then rescale their normalizations to match the observed SFR histograms. We fit only the massive end of the distribution, where the stellar mass histogram has  $>50\%$  completeness by number. The renormalized SMF-converted SFR histograms are shown in Figure 53.



**Figure 53.** Renormalized SMF-converted SFR histograms which are used to estimate the completeness of our sample when deriving the cosmic SFR density (Figure 29). See the text in Section 8.7.



## ORCID iDs

Daizhong Liu  <https://orcid.org/0000-0001-9773-7479>  
 Emanuele Daddi  <https://orcid.org/0000-0002-3331-9590>  
 Frazer Owen  <https://orcid.org/0000-0002-9819-3861>  
 Maurilio Pannella  <https://orcid.org/0000-0003-3738-3976>  
 Mark Sargent  <https://orcid.org/0000-0003-1033-9684>  
 Georgios Magdis  <https://orcid.org/0000-0002-4872-2294>  
 Yu Gao  <https://orcid.org/0000-0003-0007-2197>  
 Xinwen Shu  <https://orcid.org/0000-0002-7020-4290>  
 Tao Wang  <https://orcid.org/0000-0002-2504-2421>  
 Shuowen Jin  <https://orcid.org/0000-0002-8412-7951>

## References

- Abramson, L. E., Gladders, M. D., Dressler, A., et al. 2016, *ApJ*, **832**, 7
- Ashby, M. L. N., Willner, S. P., Fazio, G. G., et al. 2013, *ApJ*, **769**, 80
- Ashby, M. L. N., Willner, S. P., Fazio, G. G., et al. 2015, *ApJS*, **218**, 33
- Barger, A. J., Cowie, L. L., Brandt, W. N., et al. 2002, *AJ*, **124**, 1839
- Barger, A. J., Cowie, L. L., & Wang, W.-H. 2008, *ApJ*, **689**, 687
- Bertin, E., & Arnouts, S. 1996, *A&AS*, **117**, 393
- B  thermin, M., Daddi, E., Magdis, G., et al. 2015, *A&A*, **573**, A113
- B  thermin, M., Dole, H., Beelen, A., & Aussel, H. 2010a, *A&A*, **512**, A78
- B  thermin, M., Dole, H., Cousin, M., & Bavouzet, N. 2010b, *A&A*, **516**, A43
- B  thermin, M., Le Floc'h, E., Ilbert, O., et al. 2012, *A&A*, **542**, A58
- B  thermin, M., Wu, H.-Y., Lagache, G., et al. 2017, *A&A*, **607**, A89
- Bloomfield, J. K., Face, S. H. P., Guth, A. H., et al. 2016, arXiv:1612.03890
- Bourne, N., Dunlop, J. S., Merlin, E., et al. 2017, *MNRAS*, **467**, 1360
- Bouwens, R. J., Illingworth, G. D., Oesch, P. A., et al. 2012, *ApJ*, **754**, 83
- Bruzual, G., & Charlot, S. 2003, *MNRAS*, **344**, 1000
- Budav  ri, T., & Szalay, A. S. 2008, *ApJ*, **679**, 301
- Burgarella, D., Buat, V., Gruppioni, C., et al. 2013, *A&A*, **554**, A70
- Casey, C. M., Berta, S., B  thermin, M., et al. 2012, *ApJ*, **761**, 140
- Chabrier, G. 2003, *PASP*, **115**, 763
- Chapin, E. L., Berry, D. S., Gibb, A. G., et al. 2013, *MNRAS*, **430**, 2545
- Chapin, E. L., Chapman, S. C., Coppin, K. E., et al. 2011, *MNRAS*, **411**, 505
- Chapin, E. L., Pope, A., Scott, D., et al. 2009, *MNRAS*, **398**, 1793
- Chapman, S. C., Blain, A. W., Smail, I., & Ivison, R. J. 2005, *ApJ*, **622**, 772
- Chapman, S. C., Smail, I., Blain, A. W., & Ivison, R. J. 2004, *ApJ*, **614**, 671
- Ciesla, L., Boquien, M., Boselli, A., et al. 2014, *A&A*, **565**, A128
- Cohen, J. G. 2001, *AJ*, **121**, 2895
- Cohen, J. G., Cowie, L. L., Hogg, D. W., et al. 1996, *ApJL*, **471**, L5
- Cohen, J. G., Hogg, D. W., Blandford, R., et al. 2000, *ApJ*, **538**, 29
- Condon, J. J. 1992, *ARA&A*, **30**, 575
- Conselice, C. J. 2014, *ARA&A*, **52**, 291
- Cooper, M. C., Aird, J. A., Coil, A. L., et al. 2011, *ApJS*, **193**, 14
- Daddi, E., Bournaud, F., Walter, F., et al. 2010a, *ApJ*, **713**, 686
- Daddi, E., Dannerbauer, H., Elbaz, D., et al. 2008, *ApJL*, **673**, L21
- Daddi, E., Dannerbauer, H., Krips, M., et al. 2009a, *ApJL*, **695**, L176
- Daddi, E., Dannerbauer, H., Stern, D., et al. 2009b, *ApJ*, **694**, 1517
- Daddi, E., Dickinson, M., Morrison, G., et al. 2007, *ApJ*, **670**, 156
- Daddi, E., Elbaz, D., Walter, F., et al. 2010b, *ApJL*, **714**, L118
- Dahlen, T., Mobasher, B., Dickinson, M., et al. 2010, *ApJ*, **724**, 425
- Davidzon, I., Ilbert, O., Laigle, C., et al. 2017, *A&A*, **605**, A70
- Dawson, S., Stern, D., Bunker, A. J., Spinrad, H., & Dey, A. 2001, *AJ*, **122**, 598
- Delhaize, J., Smol  i  , V., Delvecchio, I., et al. 2017, *A&A*, **602**, A4
- Dickinson, M., Stern, D., Giavalisco, M., et al. 2004, *ApJL*, **600**, L99
- Dole, H., Lagache, G., & Puget, J.-L. 2003, *ApJ*, **585**, 617
- Draine, B. T. 2003, *ARA&A*, **41**, 241
- Draine, B. T., Dale, D. A., Bendo, G., et al. 2007, *ApJ*, **663**, 866
- Draine, B. T., & Li, A. 2007, *ApJ*, **657**, 810
- Dunlop, J. S., McLure, R. J., Biggs, A. D., et al. 2017, *MNRAS*, **466**, 861
- Elbaz, D., Daddi, E., Le Borgne, D., et al. 2007, *A&A*, **468**, 33
- Elbaz, D., Dickinson, M., Hwang, H. S., et al. 2011, *A&A*, **533**, A119
- Elbaz, D., Dickinson, M., Hwang, H. S., et al. 2013, *yCat*, **353**, 119
- Fazio, G. G., Hora, J. L., Allen, L. E., et al. 2004, *ApJS*, **154**, 10
- Frazer, D. T., Huynh, M. T., Chary, R., et al. 2006, *ApJL*, **647**, L9
- Frazer, D. T., Sanders, M. B., Surace, H. S., et al. 2009, *AJ*, **138**, 1261
- Gail, H.-P., Zhukovska, S. V., Hoppe, P., & Tieloff, M. 2009, *ApJ*, **698**, 1136
- Geach, J. E., Dunlop, J. S., Halpern, M., et al. 2017, *MNRAS*, **465**, 1789
- Giavalisco, M., Dickinson, M., Ferguson, H. C., et al. 2004, *ApJL*, **600**, L103
- Gladders, M. D., Oemler, A., Dressler, A., et al. 2013, *ApJ*, **770**, 64
- Gobat, R., Daddi, E., Magdis, G., et al. 2017, arXiv:1703.02207
- Grazian, A., Fontana, A., Santini, P., et al. 2015, *A&A*, **575**, A96
- Greve, T. R., Pope, A., Scott, D., et al. 2008, *MNRAS*, **389**, 1489
- Griffin, M. J., Abergel, A., Abreu, A., et al. 2010, *A&A*, **518**, L3
- Grogin, N. A., Kocevski, D. D., Faber, S. M., et al. 2011, *ApJS*, **197**, 35
- Gruppioni, C., Pozzi, F., Rodighiero, G., et al. 2013, *MNRAS*, **432**, 23
- Helou, G., Soifer, B. T., & Rowan-Robinson, M. 1985, *ApJL*, **298**, L7
- Hu, E. M., Cowie, L. L., Barger, A. J., et al. 2010, *ApJ*, **725**, 394
- Hurley, P. D., Oliver, S., Betancourt, M., et al. 2017, *MNRAS*, **464**, 885
- Ilbert, O., McCracken, H. J., Le F  vre, O., et al. 2013, *A&A*, **556**, A55
- Ivison, R. J., Magnelli, B., Ibar, E., et al. 2010, *A&A*, **518**, L31
- Kajisawa, M., Ichikawa, T., Tanaka, I., et al. 2011, *PASJ*, **63**, 379
- Kakazu, Y., Cowie, L. L., & Hu, E. M. 2007, *ApJ*, **668**, 853
- Karim, A., Schinnerer, E., Mart  nez-Sansigre, A., et al. 2011, *ApJ*, **730**, 61
- Karim, A., Swinbank, A. M., Hodge, J. A., et al. 2013, *MNRAS*, **432**, 2
- Kellermann, K. I., & Owen, F. N. 1988, in *Radio Galaxies and Quasars*, ed. K. I. Kellermann & G. L. Verschuur (Berlin: Springer), 563
- Kennicutt, R. C., Jr. 1998, *ARA&A*, **36**, 189
- Kimball, A. E., & Ivezi  , Z. 2008, *AJ*, **136**, 684
- Koekemoer, A. M., Faber, S. M., Ferguson, H. C., et al. 2011, *ApJS*, **197**, 36
- Kriek, M., Shapley, A. E., Reddy, N. A., et al. 2015, *ApJS*, **218**, 15
- Kroupa, P. 2002, *Sci*, **295**, 82
- Le Floc'h, E., Aussel, H., Ilbert, O., et al. 2009, *ApJ*, **703**, 222
- Le Floc'h, E., Papovich, C., Dole, H., et al. 2005, *ApJ*, **632**, 169
- Lee, N., Sanders, D. B., Casey, C. M., et al. 2013, *ApJ*, **778**, 131
- Lin, L., Dickinson, M., Jian, H.-Y., et al. 2012, *ApJ*, **756**, 71
- Lowenthal, J. D., Koo, D. C., Guzm  n, R., et al. 1997, *ApJ*, **481**, 673
- MacKenzie, T. P., Scott, D., & Swinbank, M. 2016, *MNRAS*, **463**, 10
- Madau, P., & Dickinson, M. 2014, *ARA&A*, **52**, 415
- Magdis, G. E., Daddi, E., B  thermin, M., et al. 2012, *ApJ*, **760**, 6
- Magdis, G. E., Elbaz, D., Dickinson, M., et al. 2011, *A&A*, **534**, A15
- Magnelli, B., Elbaz, D., Chary, R. R., et al. 2009, *A&A*, **496**, 57
- Magnelli, B., Elbaz, D., Chary, R. R., et al. 2011, *A&A*, **528**, A35
- Magnelli, B., Ivison, R. J., Lutz, D., et al. 2015, *A&A*, **573**, A45
- Magnelli, B., Popesso, P., Berta, S., et al. 2013, *A&A*, **553**, A132
- Man, A. W. S., Greve, T. R., Toft, S., et al. 2016, *ApJ*, **820**, 11
- Marvil, J., Owen, F., & Eilek, J. 2015, *AJ*, **149**, 32
- Merlin, E., Fontana, A., Ferguson, H. C., et al. 2015, *A&A*, **582**, A15
- Momcheva, I. G., Brammer, G. B., van Dokkum, P. G., et al. 2016, *ApJS*, **225**, 27
- Morrison, G. E., Owen, F. N., Dickinson, M., Ivison, R. J., & Ibar, E. 2010, *ApJS*, **188**, 178
- Mullaney, J. R., Alexander, D. M., Goulding, A. D., & Hickox, R. C. 2011, *MNRAS*, **414**, 1082
- Murphy, E. J., Chary, R.-R., Alexander, D. M., et al. 2009, *ApJ*, **698**, 1380
- Muzzin, A., Marchesini, D., Stefanon, M., et al. 2013, *ApJ*, **777**, 18
- Noeske, K. G., Weiner, B. J., Faber, S. M., et al. 2007, *ApJL*, **660**, L43
- Novak, M., Smol  i  , V., Delhaize, J., et al. 2017, *A&A*, **602**, A5
- Pannella, M., Carilli, C. L., Daddi, E., et al. 2009, *ApJL*, **698**, L116
- Pannella, M., Elbaz, D., Daddi, E., et al. 2015, *ApJ*, **807**, 141
- Pearson, W. J., Wang, L., van der Tak, F. F. S., et al. 2017, *A&A*, **603**, A102
- Peng, C. Y., Ho, L. C., Impey, C. D., & Rix, H.-W. 2002, *AJ*, **124**, 266
- Peng, C. Y., Ho, L. C., Impey, C. D., & Rix, H.-W. 2010, *AJ*, **139**, 2097
- Penner, K., Pope, A., Chapin, E. L., et al. 2011, *MNRAS*, **410**, 2749
- Perera, T. A., Chapin, E. L., Austermann, J. E., et al. 2008, *MNRAS*, **391**, 1227
- Phillips, A. C., Guzm  n, R., Gallego, J., et al. 1997, *ApJ*, **489**, 543
- Pilbratt, G. L., Riedinger, J. R., Passvogel, T., et al. 2010, *A&A*, **518**, L1
- Poglitich, A., Waelkens, C., Geis, N., et al. 2010, *A&A*, **518**, L2
- Pope, A., Chary, R.-R., Alexander, D. M., et al. 2008, *ApJ*, **675**, 1171
- Pope, A., Scott, D., Dickinson, M., et al. 2006, *MNRAS*, **370**, 1185
- Popesso, P., Magnelli, B., Buttigione, S., et al. 2012, arXiv:1211.4257
- Pozzetti, L., Cimatti, A., Zamorani, G., et al. 2003, *A&A*, **402**, 837
- Press, W. H., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. P. 1992, *Numerical Recipes in C: The Art of Scientific Computing* (2nd ed.; Cambridge: Cambridge Univ. Press)
- Reddy, N. A., Steidel, C. C., Erb, D. K., Shapley, A. E., & Pettini, M. 2006, *ApJ*, **653**, 1004
- Riechers, D. A., Carilli, L. C., Walter, F., et al. 2011, *ApJL*, **733**, L11
- Rieke, G. H., Young, E. T., Engelbracht, C. W., et al. 2004, *ApJS*, **154**, 25
- Rodighiero, G., Cimatti, A., Gruppioni, C., et al. 2010, *A&A*, **518**, L25
- Rodighiero, G., Daddi, E., Baronchelli, I., et al. 2011, *ApJL*, **739**, L40
- Roseboom, I. G., Ivison, R. J., Greve, T. R., et al. 2012, *MNRAS*, **419**, 2758
- Roseboom, I. G., Oliver, S. J., Kunz, M., et al. 2010, *MNRAS*, **409**, 48
- Rowan-Robinson, M., Oliver, S., Wang, L., et al. 2016, *MNRAS*, **461**, 1100
- Safarzadeh, M., Ferguson, H. C., Lu, Y., Inami, H., & Somerville, R. S. 2015, *ApJ*, **798**, 91

- Salpeter, E. E. 1955, [ApJ](#), **121**, 161
- Sargent, M. T., Daddi, E., Béthermin, M., et al. 2014, [ApJ](#), **793**, 19
- Sargent, M. T., Schinnerer, E., Murphy, E., et al. 2010, [ApJL](#), **714**, L190
- Saunders, W., Rowan-Robinson, M., Lawrence, A., et al. 1990, [MNRAS](#), **242**, 318
- Schmidt, M. 1968, [ApJ](#), **151**, 393
- Schreiber, C., Pannella, M., Elbaz, D., et al. 2015, [A&A](#), **575**, A74
- Scudder, J. M., Oliver, S., Hurley, P. D., et al. 2016, [MNRAS](#), **460**, 1119
- Shu, X. W., Elbaz, D., Bourne, N., et al. 2016, [ApJS](#), **222**, 4
- Skelton, R. E., Whitaker, K. E., Momcheva, I. G., et al. 2014, [ApJS](#), **214**, 24
- Song, M., Finkelstein, S. L., Ashby, M. L. N., et al. 2016, [ApJ](#), **825**, 5
- Stark, D. P., Ellis, R. S., Bunker, A., et al. 2009, [ApJ](#), **697**, 1493
- Stark, D. P., Ellis, R. S., & Ouchi, M. 2011, [ApJL](#), **728**, L2
- Steidel, C. C., Adelberger, K. L., Giavalisco, M., Dickinson, M., & Pettini, M. 1999, [ApJ](#), **519**, 1
- Steidel, C. C., Adelberger, K. L., Shapley, A. E., et al. 2003, [ApJ](#), **592**, 728
- Steidel, C. C., Giavalisco, M., Dickinson, M., & Adelberger, K. L. 1996, [AJ](#), **112**, 352
- Strolger, L.-G., Riess, A. G., Dahlen, T., et al. 2004, [ApJ](#), **613**, 200
- Strolger, L.-G., Riess, A. G., Dahlen, T., et al. 2005, [ApJ](#), **635**, 1370
- Tan, Q., Daddi, E., Magdis, G., et al. 2014, [A&A](#), **569**, A98
- Treu, T., Ellis, R. S., Liao, T. X., & van Dokkum, P. G. 2005, [ApJL](#), **622**, L5
- Vanzella, E., Giavalisco, M., Dickinson, M., et al. 2009, [ApJ](#), **695**, 1163
- Walter, F., Decarli, R., Carilli, C., et al. 2012, [Natur](#), **486**, 233
- Wang, T., Elbaz, D., Schreiber, C., et al. 2016, [ApJ](#), **816**, 84
- Werner, M. W., Roellig, T. L., Low, F. J., et al. 2004, [ApJS](#), **154**, 1
- Wirth, G. D., Trump, J. R., Barro, G., et al. 2015, [AJ](#), **150**, 153
- Wirth, G. D., Willmer, C. N. A., Amico, P., et al. 2004, [AJ](#), **127**, 3121
- Yan, H., Stefanon, M., Ma, Z., et al. 2014, [ApJS](#), **213**, 2
- Yoshikawa, T., Akiyama, M., Kajisawa, M., et al. 2010, [ApJ](#), **718**, 112
- Zhukovska, S., & Henning, T. 2013, [A&A](#), **555**, A99